الجمهورية العربية السورية

جامعة البعث ـ كلية العلوم

قسم الرياضيات ــ شعبة الانفورماتيك

استخدام النظم الخبيرة الاستقرائية في التنبؤ

دراسة أعدت لنيل درجة الماجستير في المعلوماتية

إشراف

د. ناصر أبو صالح

مدرس في كلية الهندسة المعلوماتية

جامعة البعث

اختصاص: ذكاء صنعي- شبكات عصبونية

د. عبد الله العمر

مدرس في كلية الهندسة المعلوماتية

جامعة دمشق

اختصاص: ذكاء صنعي- نظم خبيرة

تقديم الطالبة

روضة رياض شاهين

جامعة البعث ـ كلية العلوم

قسم الرياضيات ـ شعبة الإنفورماتيك

٢٠١١ م- ١٤٣٢ هـ

Syrian Arabic Republic
Albaath University
Science College
Math depart-Informatic

Using Inductive Expert Systems in Prediction

Master Degree in Informatic

Presented by
Rawda Shaheen

2011-1432

جدول الخطأ والصواب

رقم الصفحة	رقم السطر	الخطأ	الصواب
٢٣	السطر قبل الأخير	كونيلان	كوينلان
٣١	السطر التاسع عشر	فعالياً	فعلياً
٣٢	السطر الخامس	التي فقط جزء منها	جزء منها
٤٩	السطر الخامس	البشر يون	البشريون
٨١	السطر الرابع	السب	السي
٨١	السطر الرابع	اكشاف	اكتشاف
١١٨	السطر الخامس عشر	غي	في

ملخص

عرفت عملية استخراج المعرفة منذ وقت طويل بأنها المهمة الأكثر تعقيداً واستهلاكاً للزمن في تطوير النظم المبنية على المعرفة (النظم الخبيرة). وسميت بعنق الزجاجة في تطوير هذه النظم.

ولقد تغلب النجاح في اكتساب المعرفة ألياً باستخدام أساليب تعلم الآلة والتنقيب بالمعطيات إلى حد ما على هذه المهمة. و بالتالي وجدت تطبيقات في مجالات مختلفة.

ولقد طبقت أساليب تعلم الآلة والتنقيب بالمعطيات بنجاح في مجالات عديدة خاصة بمشاكل العالم الحقيقي. ولقد تلقى تطبيق أساليب تعلم الآلة والتنقيب بالمعطيات على التنبؤ بالطقس انتباهاً قليلاُ. وكان أحد الأسباب لهذا الصعوبة في الحصول على مجموعات معطيات ملائمة (مناسبة) للتنبؤ بالطقس .

اقترحنا في هذا البحث طريقة جديدة لتوليد الآلي لقواعد التنبؤ بالطقس من مجموعة معطاة من أمثلة التدريب لتطوير نظام مبني على المعرفة (نظام خبير) استقرائي للتنبؤ بطقس المدن الجمهورية العربية السورية في كل الفصول. وبنيت الطريقة في الواقع على (١) طريقة الاستقراء الموجه بالصفة بعد دراسة تأثير عناصر الطقس على بعضها البعض باستخدام نموذج تحليل الانحدار و (٢) تمديد نموذج المجموعات التقريبية . وكانت السمة البارزة الأولى لهذه الطريقة المقترحة أنها تأخذ بعين الاعتبار تأثير العناصر على بعضها البعض . وكانت السمة البارزة الثانية أنها تستفيد من المعلومات الإحصائية الكامنة في المعطيات لمعالجة عينات التدريب غير الكاملة والغامضة .

وفي إنجاز هدفنا، تتألف منهجية البحث المتبعة في هذه الأطروحة من عدة خطوات . عرضنا في الخطوة الأولى المشكلة وحللنا المجال . حيث في هذه المرحلة، عرضنا عناصر المناخ والطرق المستخدمة في التنبؤ بالطقس . ومن ثم عرضنا المحرضات (البواعث) للتنبؤ بالطقس ، و الحالة الفنية للذكاء الصنعي في التنبؤ بالطقس ، ولماذا نادراً ما يستخدم علماء الأرصاد الجوية الذكاء الصنعي، ولماذا يحتاج علماء الأرصاد الجوية إلى نظم دعم القرار، والهدف من الدراسة.

وفي الخطوة الثانية ، عرضنا مراجعة للأدب. حيث في هذه المرحلة ، راجعنا الأدب ذا الصلة بالنظم المبنية على المعرفة (النظم الخبيرة) الاستقرائية.

وفي الخطوة الثالثة ، وصفنا عملية "اكتشاف أنماط الطقس العامة في قواعد المعطيات الطقسية العلاقية الكبيرة ". حيث في هذه المرحلة ، طورنا خوارزمية جديدة ، دعيت "خوارزمية استقراء نموذج تحليل الانحدار الموجه بالصفة " لاكتشاف أنماط الطقس العامة .

وفي الخطوة الرابعة ، وصفنا عملية " اكتشاف القواعد من المعطيات للتنبؤ بالطقس باستخدام نظرية المجموعات التقريبية ". حيث في هذه المرحلة ، اقترحنا طريقة جديدة لتوليد قواعد المعرفة .

وفي الخطوة الخامسة ، نفذنا واختبرنا طريقتنا المقترحة لتوليد قواعد المعرفة في تطبيق مناخي . حيث يتعامل مع درجات الحرارة ، وكميات السحب ، والرطوبة النسبية ، وسرع الرياح، والضغوط الجوية ، وضغوط بخار الماء.

وفي الخطوة السادسة طورنا نظام مبني على المعرفة (نظام خبير) استقرائي فريد للتنبــؤ بــالطقس، دعي IESP ، باستخدام لغة C# المرئية . حيث هذا النظام هو نظام مبني على الاستقراء يستخدم نمــوذج تحليل الانحدار، ونظرية المجموعات التقريبية الموجهة بالصفة لاكتشاف أنماط الطقس العامة وقواعد التنبؤ بالطقس من قواعد المعطيات الطقسية.

وفي الخطوة الأخيرة ، اختبرنا النظام باستخدام حالات اختبارية حقيقية من السجلات الإحصائية.

Abstract

Extracting knowledge from a human expert has long been recognized as the most time-consuming and complicated task in the development of a knowledge-based system (expert system) and is termed the bottleneck of knowledge-based system development. The success in acquiring knowledge automatically by using machine learning and data mining techniques, has to a certain extent overcome this, and hence, has found applications in many fields. Machine learning and data mining techniques have been applied successfully to several real-world problem domains. The application of machine learning and data mining techniques to weather forecasting has received only little attention. One reason for this has been the difficulty in obtaining suitable weather forecasting data sets. In this research, We have suggested a new method to automatically generate weather forecasting rules from a given set of training examples for the development of an induction knowledge-based system to predict the weather of Syrian Cities in all seasons. The method is in fact based on (a) an attribute-oriented induction method after studying the effect of weather elements on each other using a regression analysis model and (b) an extension of the rough set model. The first salient feature of the proposed method is that it takes into consideration the effect of elements on each other. The second salient feature of this method is that it makes use of the statistical information inherent in the data to handle incomplete and ambiguous training samples, i.e., example sets.

In achieving our main goal, the research methodology followed in this thesis consists of several steps. In the first step, we defined the problem and analyzed the domain by using a specialized manual of weather forecasting to recognize climatic elements and methods used in weather forecasting. We then presented motivations for weather forecasting, the state of the art of AI in weather forecasting, why meteorologists have rarely used artificial intelligence, why meteorologists need decision support systems, and the aim of study. In the second step, we presented a review of literature where, in this step, we reviewed the relevant literature for

induction knowledge-based systems. In the third step, we described the "discovery of general weather patterns in large relational weather databases" process, where, in this step, we developed a new algorithm, called an attribute-oriented regression analysis model induction, for discovering general weather patterns. In the fourth step, we described the "discovering rules from data for weather forecasting using rough set theory" process, where, in this step, we proposed a new method for generating knowledge rules. In the fifth step, we implemented and tested our proposed method for generating knowledge rules in meteorological application: the application deals with temperatures, cloud quantities, humidity, wind speeds, air pressures, and vapor pressures. In the sixth step, we developed a unique induction knowledge-based system for weather forecasting, called IESP, using Visual C# where this system is an induction-based system that employs a regression analysis model and an attribute-oriented rough set theory to discover weather patterns and forecasting rules from weather databases. In the last step, we tested the system using real test cases from historical records.

الفصل الأول

مقدمة

١–١ مقدمة

الطقس هو واحد من أكثر الأمور الذي يؤثر في سلوكنا اليومي، لأننا مضطرين لتكييف أنفسنا تبعاً لحالتــه ابتداءً من ارتداءنا للملابس وانتهاءً بتخطيطنا لاستراتيجيات عملنا. كما أن حالات الطقس المفاجئة قد تؤدي إلى أضرار كبيرة في أعمالنا، لذلك نحتاج إلى أن نكون على دراية مبكرة بهذه الحالات الطقسية المفاجئــة لأخذ الحيطة والحذر وذلك عن طريق أجهزة وبرامج التنبؤ للكشف المبكر عن الظواهر الطقسية الخطيــرة وبالتالي توجد اهتمامات كبيرة بشكل عام بالمعلومات الطقسية حيث أصبح من الأمر العادي أن ننصت إلــى الأخبار الطقسية المحلية اليومية المتوقعة بدرجة كبيرة من الاهتمام. الأمر الذي دفع الكثير من الباحثين إلى الاهتمام بمجال الطقس وحثهم على تطوير أنظمة التنبؤ الطقسية القادرة على الكشــف والتنبــؤ بــالظواهر والتقلبات الطقسية باستخدام التكنولوجية العلمية [1].

في هذا الفصل، نعطي عرضاً عاماً عن التنبؤ بالطقس. وينظم هذا الفصل كما يلي: يقــدم المقطــع ١–٢ باختصار التنبؤ بالطقس. ويصف المقطع ١–٣ العناصر المؤثرة في التنبؤ بالطقس. ويصف المقطــع ١–٤ باختصار طرق التنبؤ بالطقس. ويقدم المقطع ١–٥ دوافع التنبؤ بالطقس. ويركز المقطع ١–٦ علــى الحالة الفنية للذكاء الصنعي في التنبؤ بالطقس. ويقدم المقطع ١–٧ لماذا علماء الأرصاد نادراً ما يستخدمون الذكاء الصنعي في التنبؤ بالطقس. ويعرض المقطع ١–٨ لماذا يحتاج علماء الأرصاد لنظم دعــم القــرار. ويقدم المقطع ١–٩ الهدف من الدراسة. وأخيراً يصف المقطع ١–١٠ هيكل الأطروحة.

١–٢ التنبؤ بالطقس

يعتبر التنبؤ بالطقس إحدى أهم المشاكل العلمية في وقتنا الحاضر. ولهذه المشكلة تاريخ طويل. حيث يكــون من الممتع جداً أن نعرف ماذا سيكون الطقس في الوقت القادم. وبالطبع، تكون المشكلة تحديد أهم العناصــر المؤثرة في التنبؤ بالطقس وتوضيح ما هي أفضل طرقه المستخدمة.

يكون التنبؤ بدون شك الحافز الأكبر لدراسة الطقس والمناخ. ويكون التنبؤ بالطقس لكثير من الناس مرادفاً لعلم الأرصاد الجوية [٢]. ويعتبر التنبؤ بالطقس من قبل علماء الأرصاد علم وفن على حد ســواء. ويعتمد التنبؤ بالطقس على الأساليب الموضوعية بناء" على عقود من البحث. كما يعتمد على الأساليب غير الموضوعية والرأي بناء" على الخبرة الشخصية والقواعد المحلية والممارسات العملية. وتكــون الأســاليب الموضوعية عالمية، في حين تكون بأساليب غير الموضوعية محلية. وتســتخدم الأســاليب الموضــوعية بصورة متسقة وتكون قابلة للنقل، في حين تستخدم الأساليب غير الموضوعية بصورة غير متسقة: تختلــف

الأساليب غير الموضوعية من شخص إلى آخر ومن وقت إلى آخر، ومن مكان إلى آخر. ويشير التنبؤ بالطقس إلى العبارة الدقيقة للحالات المناخية المتوقعة لمكان، أو منطقة، أو مسار معين خلال فترة زمنية معينة [٢]. وتصنف التنبؤات بالطقس غالباً إلى الأنواع الثلاثة التالية على أساس الفترة المغطاة بواسطة التنبؤات:

١- تنبؤات لفترة قصيرة لجزء أو لكامل فترة ٢٤ ساعة مع نظرة مستقبلية عن ٢٤ ساعة القادمة.
٢- تنبؤات متوسطة المدى وتكون لفترة يومين أو ٥ أيام قدماً.
٣- تنبؤات طويلة المدى وتكون لفترة أطول من ٥ أيام قدماً (أي، شهر أو فصل).

١-٣ العناصر المؤثرة في التنبؤ بالطقس

يتأثر التنبؤ بالطقس بعدة عوامل (عناصر). من بين العناصر الأخرى درجة الحرارة، وسرع الرياح واتجاهها، ونوع السحب وكميتها، ومدة سطوع الشمس، والرطوبة الجوية، والضغط الجوي، وضغط بخار الماء، والتذبذب الجنوبي/النينو، وتذبذب المحيط الشمالي. سيركز تنبؤنا فقط على بعض عناصر الطقس مثل: درجة الحرارة، وسرعة الرياح، وكميات السحب، والرطوبة النسبية والضغط الجوي، وضغط بخار الماء.

١-٣-١ درجة الحرارة

درجة الحرارة Temperature هي متوسط الطاقة الحركية للجزيئات والأداة لقياس درجة الحرارة هي الثيرمومتر المئوي بين صفرْ-١٠٠ ْ " C " Celsius. وتعد الحرارة من أهم عناصر المناخ فهي ترتبط بكل عناصره ارتباطاً وثيقاً [٥٨]. ودرجة الحرارة الجوية هي الكمية الفيزيائية التي تشير إلى سخونة وبرودة الهواء [٣].

١-٣-٢ الرياح

الرياح هي حركة جزيئات الهواء أفقياً، وهي كمية موجهه ذات سرعة واتجاه. فعندما يتحرك الهواء من حاله الثبات تصبح رياحاً، ويتم ذلك أفقياً بموازاة سطح الأرض من مناطق الضغط المرتفع إلى مناطق الضغط المنخفض نتيجة التغير الحراري [٥٨]. وتكون الشمس هي العامل المؤثر على تشكيل الرياح [٣]. واتجاه الرياح دائماً يعطى كاتجاه بحيث أنها تأتي منه. على سبيل المثال، عندما تظهر دوارة الريح أن الهواء يتحرك من الغرب إلى الشرق، نقول أن رياح غربية تهب [٤]. وتوجد أنواع كثيرة من الرياح وفقاً للاتجاه: الشمالية، الجنوبية، الغربية، الشرقية، الشمالية الغربية، الشمالية الشرقية، الجنوبية الشرقية. وتوجد

أدوات كثيرة لقياس اتجاه الرياح: دوارة الرياح، أنيموميتر الثلاثي الفناجين، وأيروفان. وتقاس سرعة الرياح غالباً بجهاز الانيموميتر، وهي عبارة عن المسافة التي تقطعها بين نقطتين في فترة زمنية محددة. وتوجد أنواع كثيرة للرياح وفقاً للسرعة: هادئة، معتدلة، إعصارية عنيفة، وإعصارية غير عنيفة، إلخ.

١-٣-٣ الضغط الجوي

الضغط الجوي عبارة عن وزن عمود من الهواء عند منسوب سطح البحر حتى نهاية الغلاف الجوي، يشير الضغط الجوي إلى القوة المبذولة من قبل جزيئات الهواء فوق منطقة معينة [٤]. ويمكن أن يعبر عن الضغط الجوي بالملي متر الزئبقي كما يقاس بالملي بار [٣]. وأجهزة قياس الضغط الجوي هي: ١- البارومتر الزئبقي، ٢- البارومتر المعدني. يعد الضغط الجوي من عناصر المناخ المهمة والذي يرتبط بها ارتباطاً وثيقاً حيث يعد السبب الرئيس في حركة الرياح بكل أنواعها [٥٨].

١-٣-٤ الرطوبة

الرطوبة عبارة عن شكل من أشكال الماء في الغلاف الجوي يتمثل في بخار غير مرئي. وهناك عدة عوامل تؤثر في التبخر الذي ينتج عنه الرطوبة وهي: درجة الحرارة، الرياح، الضغط الجوي،الارتفاع عن سطح البحر [٥٨]. ويوجد نوعان من قياس الرطوبة [٣] :

(أ) الرطوبة المطلقة

تعني الكمية المطلقة لبخار الماء في الهواء عند درجة حرارة معينة ويعبر عنها (g/m^3). ويدعى المحتوى الأعظمي لبخار الماء في الهواء عند درجة حرارة معينة بالرطوبة المشبعة[٥٨].

(ب) الرطوبة النسبية

الرطوبة النسبية هي نسبة كمية بخار الماء في الهواء مقارنةً بالكمية الأعظمية لبخار الماء المطلوب للإشباع عند درجة حرارة معينة [٢]. وإنها فقط شكل نسبي، لا يشير إلى القيمة المطلقة للرطوبة في الهواء. والأداة لقياس الرطوبة هي الهايغرومتر الشعري.

١-٣-٥ السحب

يعد شكل من صور بخار الماء المرئي تكثف بسبب انخفاض درجات الحرارة ووصولها إلى ما دون نقطة الندى في وجود نويات التكاثف، على شكل دقائق مائية تقاس بالميكرومتر، وتعكس السحب ما يتراوح بين

٤٠-٨٠% من الأشعة الشمسية الساقطة عليها، وتحافظ على الإشعاع الأرضي وحرارة الهواء من التبدد في الفضاء ويسقط منه الغيث [٥٨]. وتجمع السحب عادة تحت ثلاثة أصناف رئيسية وفقاً لارتفاع قواعدها: عالية، متوسطة، ومنخفضة [٣,٤].

١-٣-٦ ضغط بخار الماء

يُعرف ضغط بخار الماء بأنه الضغط المبذول من قبل بخاره عندما يكون الماء وبخار الماء في توازن ديناميكي. ويعبر عن ضغط بخار الماء بالملي بار (mb). ويوجد نوعان من ضغط بخار الماء: ضغط بخار الماء الفعلي، وضغط بخار الماء المشبع. ويشير ضغط بخار الماء الفعلي إلى محتوى بخار الماء الكلي للهواء. ويشير ضغط بخار الماء الفعلي المرتفع إلى أعداد كبيرة من جزيئات الماء. في حين يشير ضغط بخار الماء الفعلي المنخفض إلى أعداد صغيرة من جزيئات بخار الماء. ويصف ضغط بخار الماء المشبع كمية بخار الماء الضرورية لجعل الهواء مشبعاً عند أية درجة حرارة معينة. بكلمات أخرى، يكون ضغط بخار الماء المشبع الضغط الذي تبذلهُ جزيئات بخار الماء عندما يكون الهواء مشبعاً عند درجة حرارة معينة. ويعتمد ضغط بخار الماء المشبع أساساً على درجة حرارة الهواء [٤].

١-٤ طرق التنبؤ بالطقس

في أيامنا هذه، لدينا نوعان من طرق التنبؤ بالطقس [٥]. في هذا المقطع، نصف باختصار هذه الأنواع بدون الدخول بالتفاصيل.

يكون النوع الأول منها طريقة التنبؤ بالطقس الرياضي. وتستخدم هذه الطريقة بصورة واسعة في التنبؤ بالطقس. وتكون مبنية على الصيغ الرياضية التي تبني الانحدار للتنبؤ بسلوك الطقس. ويوجد نوعان من الانحدار: الانحدار البسيط، والانحدار المتعدد.

ويُبنى النوع الآخر من طرق التنبؤ بالطقس على أساليب الذكاء الصنعي، مثل النظم الخبيرة، وخوارزميات تعلم الآلة، والشبكات العصبية. وتنجز هذه الطريقة التنبؤ بالطقس على مرحلتين: تكون المرحلة الأولى تعلم الآلة، وتكون المرحلة الثانية التنبؤ.

١-٥ دوافع التنبؤ بالطقس

توجد دوافع كثيرة للتنبؤ بالطقس، ومن هذه الدوافع، الآتي:

- تُصدر التنبؤات بالطقس لحماية الأملاك والمحاصيل وإخبارنا ما يتوقع في بيئتنا الجوية. بالإضافة إلى ذلك، تكون معرفة ماذا سيكون الطقس في المستقبل حيوية بالنسبة للنشاطات البشرية.
- يُنذر التنبؤ المزارعين لأن يجنوا محاصيلهم قبل أن تصبح حقولهم ندية وبذلك يتخذون إجراءاتهم بشأن ذلك.
- يُنذر التنبؤ الملاحين الجويين على تأجيل رحلاتهم عندما يكون الطقس سيئاً جداً في السلم والحرب.
- يحب الناس أن يعرفوا الحالة الجوية مسبقاً قبل أن يذهبوا إلى أعمالهم.
- يحب الزائرين لمنطقة ما أن يعرفوا ماذا سيكون الطقس أثناء زيارتهم.
- سيتعين على منظمي الألعاب الرياضية أن يغيروا جدول أعمالهم أو موقع ألعابهم إذا لم تكن الحالات الطقسية مناسبة.

١–٦ الحالة الفنية للذكاء الصنعي في التنبؤ بالطقس

تتقدم الحالة الفنية للذكاء الصنعي باستمرار. ومسح كريستوفرسون (١٩٩٨) مؤخراً علم الأرصاد الجوية وحدد أكثر من أربعين مقالة علمية عن الذكاء الصنعي في مجال الأرصاد الجوية. في حين نفس المسح قام به كونواي منذ عشر سنوات (١٩٨٩) فحدد أربع مقالات علمية فقط. وتُستخدم الآن أنظمة التنبؤ العملياتية، التي تستخدم الذكاء الصنعي، من قبل خدمة الأرصاد الجوية الكندية (MSC) والجيش الأمريكي والأسطول الأمريكي. وفي عام ١٩٩٨، أعطت جمعية الأرصاد الجوية الأمريكية موافقتها على الذكاء الصنعي بعقد مؤتمرها الأول عنه.

وحدد كونواي (١٩٨٩) ثلاث تحديات يضعها علم الأرصاد الجوية في الذكاء الصنعي [٦]:

* الحاجة للملائمة والسرعة.

* مشاكل التعرف على الأنماط.

* المعطيات المفقودة والمتضاربة.

ويشرح (١٩٨٩) كونواي كيف يستنتج (يفكر) المتنبئون بمعطيات غير دقيقة [٦] كما يلي:

نحن كبشر لا نستنتج عادةً بالمصطلحات العددية ولكن نفضل الأفكار الأكثر غموضاً للأشياء والتي يمكن أن تكون " محتملة " أو " ممكنة "، لذلك يكون التخصيص المناسب للاحتمالات إحدى الصعوبات الرئيسية لتشفير الخبرة البشرية في شكل القواعد. وتكون كيفية التعامل بشكل أمثل مع الاستنتاج المنطقي تحت عدم التأكد موضوع بحث مستمر في مجتمع النظم الخبيرة.

وتكون القوة الدافعة وراء تطوير نظم الذكاء الصنعي في الأرصاد الجوية، الحاجة للتعامل بشكل فعال أكثر مع الدفق الكبير للمعطيات التي يعتمد عليها التنبؤ. على سبيل المثال، نحن نتلقى حوالي عشرة

ميغابايت بالثانية من المعطيات المستشعرة عن بعد من الأقمار الصناعية. وينتج أيضاً التنبؤ بالطقس العددي (MSC) كميات ضخمة من المعطيات التي تحتاج لأن تدمج مع أنواع المعلومات التنبئيـــة الأخـــرى فـــي التنبؤات. ويحتاج المتنبئون بالطقس إلى أنظمة حاسوبية محسنة وأنظمة ذكاء صنعي محسنة لكي يســـتغلوا بشكل أفضل كميات المعطيات الضخمة والمتزايدة [٤].

ويحدد العالم " كلين " مشكلتين تواجهان مطوري النظم الخبيرة للتنبؤ بالطقس. أولاً يمكن أن يكـــون من الصعب إيجاد الخبير الحقيقي (على سبيل المثال، خبيران يمكن أن يتناقضان مع بعضهما البعض). ثانياً توجد مآزق كامنة في ممارسة الطلب من الخبير لأن يصف الحالات غير العادية أو الصعبة فيمـــا يخـــص استثناء الحالات العادية. ويمكن أن تمثل المواقف غير المألوفة بواسطة محرك الاستدلال (آلية الاســـتدلال) وهذا سيحاكي الميل العرضي للمتنبئين بالطقس لأن يتنبئوا بالطقس على نحو أشمل. ويمكن أن يقود الطلب من المتنبئ لأن يصف كل المواقف الصعبة إلى رؤية معقدة غير ضرورية لعملية التنبؤ [٧].

ويقترح ماير"أن الذكاء الصنعي يمكن أن يملك دور في مساعدة المتنبئ في تفسير مخرجات النماذج العددية NWF وضبطها من أجل الحالات المحلية [٨] ".

ويدعى موشر (١٩٩٨) انه " حتى مع نماذج التنبؤ الجديدة، يمكن أن يضيف المتنبئ بالأرصاد الجوية قيمة إلى دليل NWF. ويمكن أن يقدم المتنبئ معلومات فريدة لا تكون متاحة من NWF ". وبصورة مشـــابهة، يمكن أن يُجمع الذكاء الصنعي معطيات فريدة من المصادر المكملة، مثل أرشيفات الطقس، وNWF [٩].

ويعبر كريستوفرسون (١٩٩٨) عن تفاؤله بمستقبل الذكاء الصنعي في الأرصاد الجوية عندما يختتم مسحه كما يلي [١٠]:

" تمتلك درجة التعقيد للتنبؤ بالطقس الحديث صفات عديدة تعرف لتطبيق الذكاء الصنعي. ويحتاج المتنبئـــون إلى مساعدة لكي يستفيدوا من فيض المعطيات ولكي يطوروا عملية التنبؤ الحديثة. وتقدم أســـاليب الـــذكاء الصنعي، بشكل خاص النظم الخبيرة والشبكات العصبية، الحلول لهذه المشاكل. "

ويقيد كريستوفرسون تفاؤله بوصف العقبات لقبول والاستخدام الذكاء الصنعي في التنبـــؤ بـــالطقس [١٠]. توصف هذه العقبات في المقطع ١–٧.

واستخدم كومار وآخرون (١٩٩٤) أسلوب تعلم الآلة للتعلم الاستقرائي للحصول على قواعد التنبـــؤ [١١]. وكانت قواعد التنبؤ هذه الأساس لنظام التنبؤ بسقوط المطر لمدة (٢٤) ساعة في مدينة ميلبورن فـــي استراليا. وكانت مشكلتهم إحداث تنبؤات تصنيفية لسقوط المطر خلال فترة (٢٤) ساعة في مدينة ميلبورن. وكان لديهم مجموعة معطيات لثلاثين سنة. وأنهم استخدموا ١٢٩ صفة. ٥٩ من هذه الصفات كانـــت مـــن المجالات التكهنية (التنبؤية) لـــ NWF. لذلك، هم ربطوا الدليل المناخي مع دليل NWF. ولقد اســـتخدموا برامج التعلم الاستقرائية لبناء أشجار القرار. ومُثِلَت المخرجات لبرامج التعلم كمجموعـــات مـــن القواعـــد وطُلب من المتنبئين لأن يعلقوا على هذه القواعد.

١-٧ لماذا نادراً ما يستخدم علماء الأرصاد الذكاء الصنعي ؟

بناءً على تحاور مكثف مع المطورين المتنبئين وباحثي الأرصاد الجوية العاملون في تطوير الذكاء الصنعي وعلى مسح لأكثر من أربعين نظاماً للذكاء الصنعي في الأرصاد الجوية، عرض (سرد) كريستوفرسون الأسباب المحتملة للقبول والاستخدام المحدودين للذكاء الصنعي في مكاتب التنبؤ العملياتية كما يلي [١٠]:

١- الافتقار إلى الخطط الخاصة على المستوى القومي لدمج أو توحيد هذه التقنيات في عملية التنبؤ.

٢- الافتقار إلى البيئة الحاسوبية الفردية في المجال التي لديها القوة والوصول المرن لدمج (توحيد) مجموعات المعطيات المختلفة والمعقدة وغير الثابتة. وهذا يكون على وجه الخصوص صحيحاً لأن خبراء المجال (علماء الأرصاد) نادراً ما يكونوا مبرمجي نظم، أو مدربين من الناحية الأكاديمية في الذكاء الصنعي.

٣- تكون عملية تطوير الذكاء الصنعي عملية هندسية أكثر من كونها بحثاً علمياً. وعلماء الأرصاد يكونوا مدربين على البحث العلمي وعادة يتجنبون أو قد يحتقرون العملية الهندسية.

٤- غالباً يكون الذكاء الصنعي عملية غير خطية (التغيرات الدقيقة في المدخلات تؤدي إلى تغيرات كبيرة في المخرجات). وإنه ليس مبنياً على النموذج الطبيعي لمجال المشكلة. وهذا يعوق علماء الأرصاد، الذين يريدون الحلول الخوارزمية التي تنمذج الجو.

٥- بينما تكون أساليب الذكاء الصنعي تقنيات عامة ومتنوعة، تحل معظم التطبيقات مشاكل ضيقة. ويتطلب تغيير أي شيء إعادة عمل تطويري كامل لاسترداد أي مهارة ويمكن أن يتطلب تغيرات جوهرية في تصميم النظام.

٦- تقترح الطبيعة الخاصة غالباً لحلول الذكاء الصنعي أنها تستخدم بشكل أفضل في مكاتب التنبؤ الإقليمية بدلاً من مركز التنبؤ القومي. ومن ناحية ثانية، تقليدياً، يجهز مكتب التنبؤ على نحو ضعيف للعمل مع مجموعات المعطيات التفصيلية، وهناك افتقار إلى مجموعات المعطيات المفصلة.

٧- لن يقبل الذكاء الصنعي حتى يُطور ويُعلم ويُستخدم في الجامعة وبرامج الأرصاد الجوية للكلية وفي المختبرات البحثية ووسائل التدريب.

١-٨ لماذا يحتاج علماء الأرصاد الجوية إلى نظم دعم القرار ؟

بمقابلة المتنبئون، وجد كومار وآخرون أربع أسباب لماذا يحتاج المتنبئون إلى نظم دعم القرار. ويمكن صياغة هذه الأسباب كما يلي [١١]:

١- المتنبئون يتحدون في ضبط عملهم الحالي من ناحية امتصاص وفهم وتذكار كمية المعلومات الكبيرة التي تصل في تيار مستمر. وكنتيجة، فإن المتنبئون أحياناً يرتكبون أخطاء في أحكامهم.
٢- من الصعب أن نكتشف من خلال خبرة التنبؤ كيف يمكن القيام بتنبؤات قريبة من المثالية.
٣- يعبر المتنبئون أنفسهم عن عدم التأكد في كيفية الاستخدام الأفضل لمعلومات دليل التنبؤ المتاحة. وحتى الخبراء المتمرسون لا يعرفون كيف يستخدمون بشكل أفضل معلومات الدليل.
٤- لا توضح بعض إحصائيات التأكد من التنبؤ أي تحسين في مهارات التنبؤ طوال العقود الأخيرة برغم التحسينات في جودة وكمية معلومات الدليل طوال نفس الفترة الزمنية.

١-٩ هدف الدراسة

عُرفت عملية استخراج المعرفة من الخبير البشري منذ وقت طويل بأنها المهمة الأكثر تعقيداً والأكثر استهلاكاً للزمن في تطوير النظم الخبيرة وسميت بعنق الزجاجة في تطوير هذه النظم.

وتوجد مشاكل عديدة باكتساب المعرفة من الخبير البشري التي تجعل منها مهمة صعبة فمن الممكن أن يكون الخبير غير مدركاً للمعرفة المستخدمة، أو غير قادراً عن التعبير عن هذه المعرفة.

وبسبب هذه المشاكل المرتبطة باستخراج المعرفة من الخبير البشري، هدفنا في هذا البحث التغلب على هذه المشاكل بأتمتة عملية اكتساب المعرفة بتطوير منهج جديد يعتمد على مفاهيم وأساليب من تعلم الآلة، والتنقيب بالمعطيات، ونموذج تحليل الانحدار. وسيطبق هذا المنهج على مجموعة المعطيات الطقسية لاكتساب المعرفة آلياً لتطوير نظام خبير للتنبؤ بالطقس.

١-١٠هيكل الأطروحة

تشمل هذه الأطروحة على سبعة فصول وملحق. بعد هذا الفصل التمهيدي، يعرض الفصل الثاني دراسة مرجعية مركزاً على النظم الخبيرة الاستقرائية وأهم الأساليب الشائعة لتعلم الآلة والتنقيب بالمعطيات والتي تستخدم في بناء النظم الخبيرة الاستقرائية. ويصف الفصل الثالث اكتشاف أنماط الطقس العامة في قواعد المعطيات الطقسية العلائقية الكبيرة. ويصف الفصل الرابع اكتشاف القواعد من المعطيات للتنبؤ بالطقس باستخدام نظرية المجموعات التقريبية كما يعرض التنفيذ لاختبار الخوارزميات المطورة. ويصف الفصل الخامس نظامنا الفريد للتنبؤ بالطقس المبني على الاستقراء باستخدام استقراء نموذج الانحدار الموجه بالصفة ونظرية المجموعات التقريبية. ويركز الفصل السادس على عملية اختبار النظام. ويعرض الفصل السابع نتائجنا ويصف الاتجاهات المستقبلية الممكنة لهذا البحث. ويعرض الملحق التحاور بين نظام IESP والمستخدم.

الفصل الثاني

دراسة مرجعية

٢-١ مقدمة

في البداية لا بد لنا أن نعرف ما هي المعرفة؟ المعرفة هي فهم نظري، أو عملي لموضوع معين، أو نطاق معين. كما أن المعرفة هي حاصل جمع ما هو معروف حالياً، وتكوّن المعرفة الواضحة قوة. ويسمي هؤلاء الذين لديهم المعرفة خبراء. وهم الناس الأكثر قوة، والأكثر أهمية في تنظيماتهم. ويكون لدى أي شركة ناجحة بضع خبراء من الطراز الأول على الأقل، ولا يمكن لنجاحها أن يستمر من دونهم.

من الذي يعرف أنه خبير(مهندس معرفة) بصفة عامة؟ يمكن اعتبار أي فرد خبير نطاق إذا كانت لديه معرفة عميقة (بكل من الحقائق، والقواعد) وخبرة عملية قوية في نطاق معين. ويمكن أن يكون مجال النطاق محدوداً. مثال ذلك، يمكن أن يكون لدى خبراء الآلات الكهربائية معرفة عامة فقط بالمحولات بينما يمكن أن يكون للخبراء في تسويق التأمين على الحياة فهماً محدوداً ببوليصة التأمين على العقارات. وبصفة عامة، الخبير هو فرد ماهر يمكنه أن يؤدي ما لا يستطيع الناس الآخرون تأديته.

توجد عدة أدوات لمساعدة صناع القرار لأن يتخذوا أفضل القرارات. ويشار إلى هذه الأدوات بالنظم الخبيرة.

كما نعرف، تكون القوة خلف النظام الخبير معرفته. ويمكننا أن نتتبع نجاح أو فشل النظام بنوعية وجودة معرفته. ومن ناحية ثانية، يمكن أن يكون الحصول على هذه المعرفة من الخبير مهمة صعبة لأن العملية الذهنية البشرية تكون داخلية، ويكون تمثيلها على هيئة خوارزمية أمر معقد جداً.

فيجب أن يكون مهندس المعرفة متمكناً في أساليب استخراج المعرفة التي تتطلب مهارات اتصال جيده وفهم ما لفلسفة التفاعل البشري. ويمكن أن يكون اكتساب هذه المهارات والقدرة على ممارستها بفاعلية صعباً. ويحدث تحدي أخر مع استخراج المعرفة عندما الخبير لا يكون مدركاً للمعرفة المستخدمة. ويصبح الشخص خبيراً في المجال من خلال سنوات من الخبرة في حل مشاكل المجال. ومن هذه الخبرة، يجمع غالباً الخبير المعرفة في خطوات كمية لحل المشكلة التي تسمح بحل فعال للمشكلة. وفي هذه الحالة، يمكن أن يكون من الصعب أن نترجم معرفة الخبير في الشكل الذي يمكن أن يستخدم في النظام الخبير.

وعند تطوير النظام الخبير المبني على القواعد، تواجه أحياناً صعوبة إضافية حيث الخبير لا يكون قادراً على نقل المعرفة في شكل قاعدة. وفي هذه الأمثلة، يمكن أن يقول الخبير، "حسناً، إنه من الصعب أن أشرح ولكن بالإمكان أن أعطيك بعض الأمثلة. " وعندما يحدث هذا، يجب أن يستخدم مهندس المعرفة طرق أخرى التي يكون بإمكانها اكتشاف المعرفة المخفية في الأمثلة بحيث يكون من الممكن تمثيلها في شكل قاعدة.

و هناك صعوبة أخرى تظهر في تطبيقات النظم الخبيرة وهي عدم وجود خبراء حقيقيين. لنأخذ على سبيل المثال، مجال التنبؤ بالطقس. بالرغم من وجود أفراد واسعي المعرفة يكون بإمكانهم التنبؤ بحدث ما، إلا إنهم عادة يعتمدون على الأحداث الماضية لتساعدهم في تنبؤاتهم. وبناء على ذلك، تكمن معرفة المجال في الأمثلة الماضية وليس مع الخبرة البشرية. ولذلك، في هذه التطبيقات، يجب أن يستخدم مهندس المعرفة مرة أخرى بعض الأساليب التي يمكن أن تكتشف المعرفة المحتواة ضمن الأمثلة.

وبسب هذه المشاكل المرتبطة باستخراج المعرفة من خبير المجال (النطاق)، تكون الأساليب التي يمكن أن تأتمت عملية اكتساب المعرفة من خلال تعلم الآلة منشودة أو مطلوبة جدا.

يراجع هذا الفصل، النظام الخبير في المقطع ٢-٢، كما يراجع أهمية التنقيب في قواعد المعطيات في المقطع ٢-٣، وإحدى الطرق المعروفة بالاستقراء في المقطع ٢-٤، والخطوات الرئيسية في تطوير النظام الخبير الاستقرائي في المقطع ٢-٥، ومزايا الاستقراء في المقطع ٢-٦، ومساوئ الاستقراء في المقطع ٢-٧، والنظم الخبيرة الطقسية المطورة من خلال الاستقراء في المقطع ٢-٨.

٢-٢ النظام الخبير

النظم الخبيرة، أحد أقوى فروع الذكاء الاصطناعي الذي يعتبر بدوره أقوى فروع علم الحاسب الآلي، وتتميز بأنها المجال الذي يتم فيه تمثيل المعرفة والتحكم في البحث داخل قواعد المعرفة. بمجرد أن يقدم الخبير البشري المعرفة يمكننا أن ندخلها في الحاسب. ونتوقع من الحاسب أن يعمل كمساعد ذكي في نطاق خبرة محدد معين، أو يحل المشكلة التي تحتاج في حلها إلى خبير. ونريد أن يكون الحاسب قادراً أيضا على عمل تكامل للمعرفة الجديدة، ويبين معرفته في صورة سهلة القراءة والفهم، ويتعامل مع جمل بسيطة بلغة طبيعية بدلاً من لغة البرمجة الصناعية. أخيراً نريد من الحاسب أن يشر كيف وصل إلى استنتاج معين. وبكلمات أخرى علينا أن نبني نظام خبير، برنامج حاسب يمكنه العمل على مستوى الخبير البشري في مجال نطاق ضيق [٢٧].

٢-٢-١ تعريف النظام الخبير

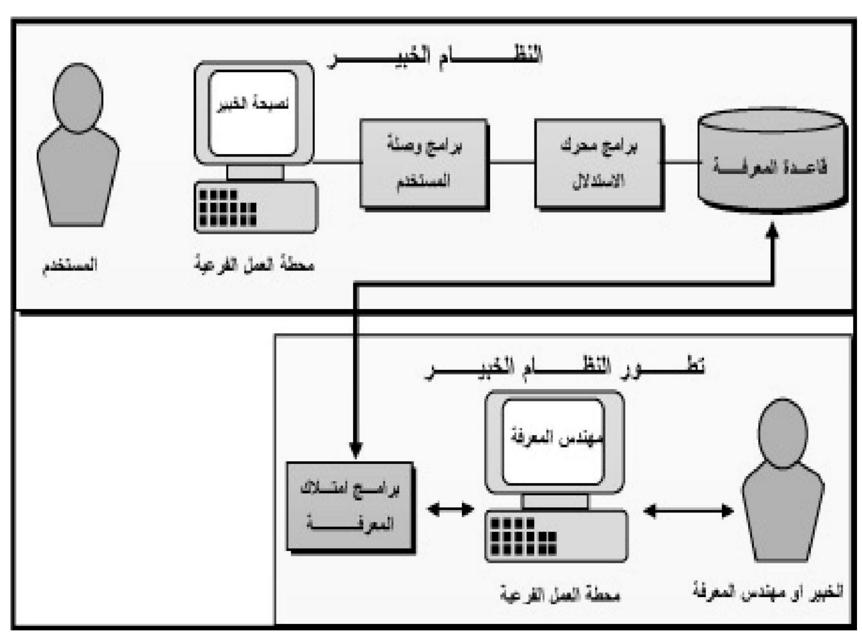
هو برامج يحاكي أداء الخبير البشري في مجال خبرة معين، وذلك عن طريق تجميع واستخدام معلومات وخبرة خبير أو أكثر في مجال معين. باختصار تم إيجاد هذا النظام من أجل استخلاص خبرات الخبراء – وخصوصاً في التخصصات النادرة – وضمها في نظام خبير يحل محل الإنسان ويساعد في نقل هذه الخبرات لأناس آخرين بالإضافة إلى قدرته على حل المشكلات بطريقة أسرع من الخبير البشري.

كما عرف الباحث الدكتور إدوارد فينجن باوم النظام الخبير بأنه البرنامج الذكي الذي يستخدم القواعد المأخوذة من الخبرة الإنسانية على هيئة شروط ونتائج في مجال معين واستخدام طرق الاشتقاق والاستدلال

لاستخراج واستنتاج النتائج المعللة بالأسباب والناتجة عن تطابق هذه الشروط أو النتائج مع شرط أو نتيجة ما والخاصة بمشكلة معينة يراد إيجاد حل لها.

وتكون النظم الخبيرة الأكثر شيوعاً مبنية على القواعد. وأصبحت أغشية النظم الخبيرة شائعة الاستخدام خاصة في تطوير النظم المبنية على القواعد. وميزتها الأساسية هي أن باني النظام يمكنه التركيز على المعرفة نفسها بدلا من التركيز على تعلم لغة البرمجة. وغشاء النظام الخبير هو نظام خبير محذوف منه المعرفة. لذلك فكل ما على المستخدم أن يعمله هو إضافة المعرفة في صورة قواعد، وتوفير البيانات المناسبة لحل المشكلة [٢٧، ٢٨].

٢-٢-٢ مكونات النظام الخبير



الشكل (٢-١) مكونات النظام الخبير

يتكون من قسمين:

القسم الأول: والذي يتم فيه بناء قاعدة المعرفة (Knowledge Base) وتحتوي على كل الحقائق والقواعد والعلاقات التي تمثل المعرفة (معلومات وعمل الخبراء)، إذن قاعدة المعرفة هي مجموعه من الحقائق والقواعد التي توضع في شكل جمل برمجيه يمكن أن تكتب بإحدى لغات برمجه الذكاء الاصطناعي مثل (lisp_ prolog) لذلك فان هذه المعرفة تسمى الذاكرة المؤقتة للنظام الخبير أو ذاكره العمل Working Memory.ويتم وذلك بالتسلسل التالي [٢٩]:

١- الخبير(Expert) أو مجموعة الخبراء في المجال المعين المطلوب الحصول على النظام الخبير فيه، حيث يقوم بإعطاء كل ما جمعوه من خبرة بأدق التفاصيل إلى مهندس المعرفة، لذلك يجب أن يكون

الخبير قادراً على توصيل معرفته، ومرحباً بالمشاركة في تطوير النظام الخبير، وملتزماً بتوجيه الوقت اللازم للمشروع. ويعد خبير النطاق الفرد الأكثر أهمية في فريق تطوير النظام الخبير.

٢- مهندس المعرفة (Knowledge Engineer) : يتولى وضع الخبرة في شكل قواعد للتضمين الشرطي متضمنة للشروط والنتائج لهذه الشروط وكذلك تطبيق التقنيات المختلفة. لذلك يجب أن يكون مهندس المعرفة قادراً على تصميم النظام الخبير، وبنائه واختباره. ويكون مسؤولاً عن اختيار المهمة المناسبة للنظام الخبير. ومن خلال التداخل مع الخبير يحدد مهندس المعرفة طرق التفكير التي يستخدمها الخبير في التعامل مع الحقائق، والقواعد ويحدد كيف يمثلها في النظام الخبير. لذلك يكون مهندس المعرفة ملتزماً بالمشروع منذ مرحلة التصميم الابتدائي وحتى التسليم النهائي للنظام الخبير، وحتى بعد إتمام المشروع فإنه يظل مشمولاً في صيانة النظام.

٣- المبرمج Programmer: هو الفرد المسؤول عن البرمجة الفعلية، ويصف معرفة النطاق بالنسبة إلى ما يمكن أن يفهمه الحاسب. ويحتاج المبرمج إلى أن يكون لديه مهارات في البرمجة.

القسم الثاني: يتم فيه طرح الأسئلة من قبل المستخدم والحصول على النتائج أو النصيحة من النظام الخبير والذي يتم بالتسلسل التالي [٢٩]:

١- المستخدم النهائي (End-User): العديم الخبرة أو ذو الخبرة المحدودة والذي يريد أن يحصل على المعرفة الحقيقية والخبرة المكتسبة لمشكلة ما في هذا المجال.

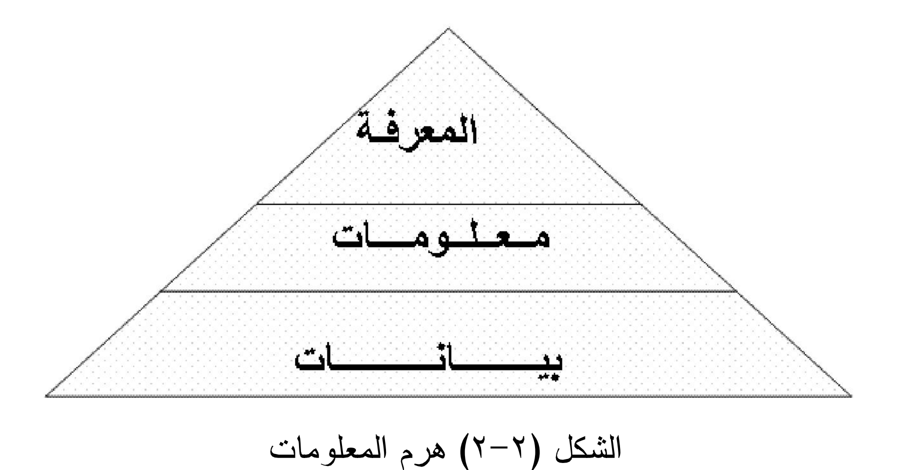
٢- واجهة المستخدم (User Interface): التي تسمح بتبادل المعرفة بين المستخدم وقاعدة المعرفة والتي تجعل سؤال المستخدم مفهوماً من قبل قاعدة المعرفة.

٣- آلية التحكم (Control Mechanism): هي التي تقوم بالتحكم الداخلي في محرك الاستدلال وقاعدة المعرفة للإجابة على السؤال المطلوب من قبل المستخدم.

٤- محرك الاستدلال (Inference Engine) : الذي يقوم بعمل يشبه عمل المحرك وذلك بتوجيه البحث في الاتجاهات المختلفة بقاعدة المعرفة إلى أن تتم الإجابة على سؤال المستخدم. وهي تقوم بالتفكير حتى يصل النظام الخبير إلى حل. أي تقوم بربط القواعد الموجودة في قاعدة المعرفة بالحقائق التي توفرها قاعدة البيانات.

حيث أن قاعدة المعرفة تحتوي على معرفة النطاق المفيدة في حل المشكلة. وفي النظام الخبير المبني على القواعد تمثل المعرفة كمجموعة من القواعد. وتحدد كل قاعدة علاقة أو توصية أو توجيه، وتأخذ هيكل عبارة IF التالي:

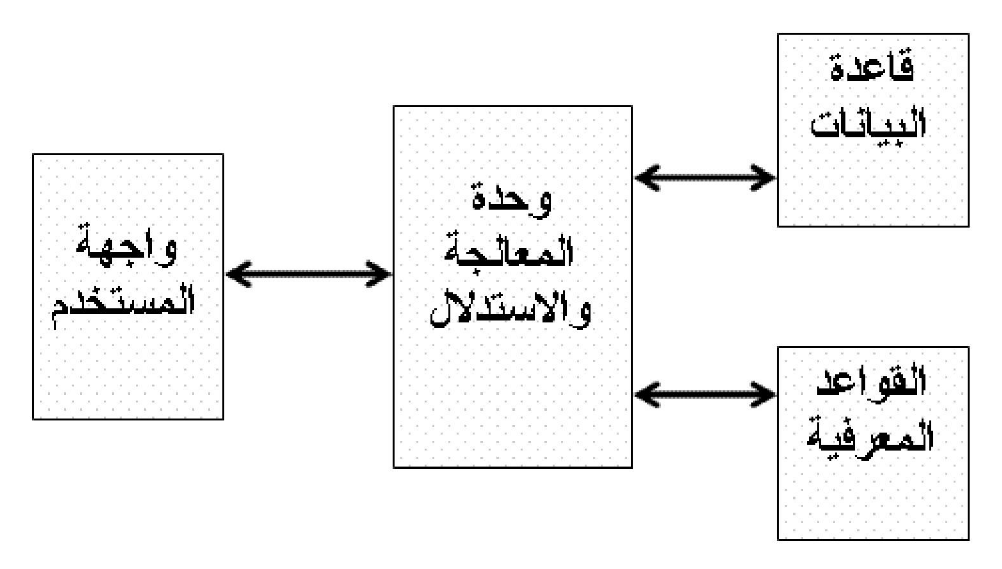
IF(condition)then (action)



الشكل (٢-٢) هرم المعلومات

و عندما يتحقق جزء القاعدة يقال أن القاعدة قد أطلقت فعل وينفذ جزء الإجراء.

بينما تشمل قاعدة البيانات مجموعة من الحقائق المستخدمة في التوفيق مقابل أجزاء IF(condition) للقواعد المخزنة في قاعدة المعرفة.



الشكل (٢-٣) أجزاء النظام الخبير

٢-٢-٣ مراحل بناء النظم الخبيرة

تمر عملية بناء النظم الخبيرة بالخطوات الآتية [٢٧]:

١- جمع الحقائق العلمية والتاريخية المؤكدة كلها عن موضوع الخبرة.

٢- مناقشة الإنسان الخبير فيما اكتسب من معارف مؤكدة ومراجعتها، وتحديد ما يجب تضمينه للنظام الخبير، وتكوين المفاهيم والعلاقات بينها وتوضيح ذلك بالرسوم والأشكال.

٣- وضع قواعد العمل المنطقية التي تتمثل في استخدام الجملة الشرطية "إذا... ، إذن."

٤- الدمج بين الحقائق والقواعد، وإدخالها إلى الحاسوب عبر قاعدة بيانات تسمى قواعد المعرفة وفق ضوابط محددة.

٥- وضع آلة للاستدلال التي تطبق القواعد على جميع الحقائق التي يغذى الجهاز بها بالاستناد إلى قاعدة المعرفة.

٦- اختبار النظام باستشارته عن مسائل حقيقية وعرض الأجوبة (الاستشارة المقترحة) على خبراء حقل المعرفة ومناقشتهم فيها، وتعديل قاعدة المعرفة إذا لزم الأمر، وهكذا حتى يصبح النظام جاهزا للعمل.

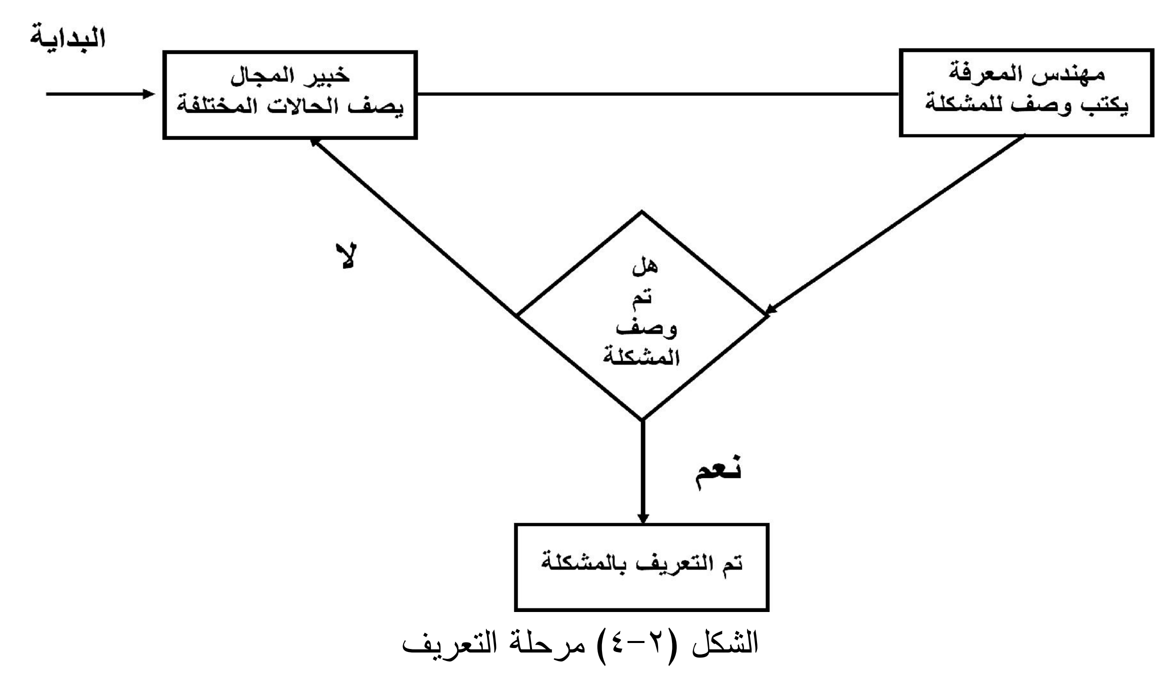
٧- تغذية قاعدة المعرفة بجميع التطورات والمستجدات العلمية والتجريبية على حقائق حقل الخبرة.

٨- تدريب الأشخاص الذين سيتعاملون مع النظام.

٢-٢-٤ بناء النظم الخبيرة

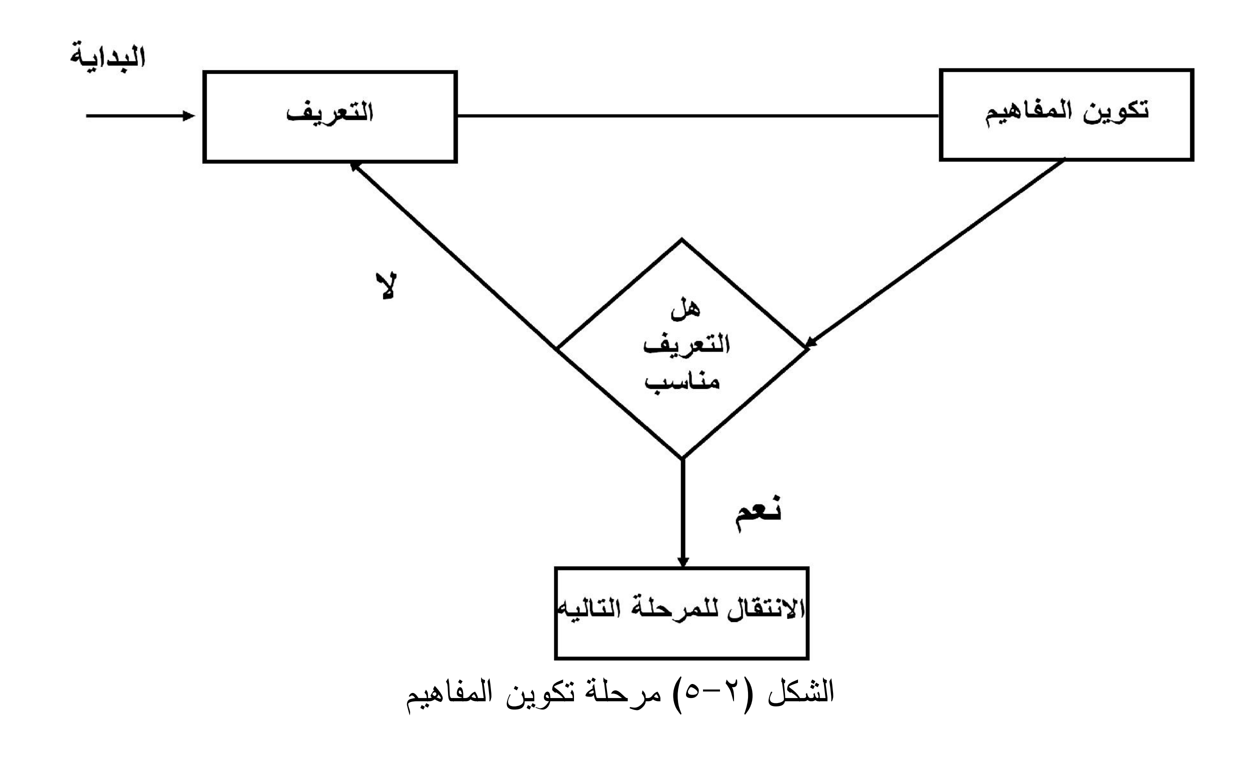
عندما نقوم ببناء النظم الخبيرة عادة ما يتم إسناد العمل إلى فريقين: مهندسوا المعرفة، وخبراء المجال، و يتم بناء النظام الخبير من خلال خمسة مراحل حيث يشترك خبير المجال مع مهندس المعرفة في المرحلة الأولى والثانية أما المراحل الأخرى فهي خاصة بمهندس المعرفة. والمراحل هي:

١. التعريف Identification: يقوم مهندس المعرفة بجمع اكبر قدر من المعلومات من خبير المجال حول المشكلة المعينة والتي قد يحتاج فيها مهندس المعرفة إلي عقد عدة جلسات مع خبير المجال حيث سيقوم خبير المجال بشرح معرفته في هذا الميدان والطرق التي يتبعها في حل المسائل وقد يقدم شرحه بطريقه غير منظمه لأنها ربما المرة الأولى التي يطلب منه القيام بذلك، ويمكن توضيح ذلك بالشكل (٢-٤)



الشكل (٢-٤) مرحلة التعريف

٢- تكوين المفاهيم Conceptualization: في هذه المرحلة يتم تحليل المشكلة بدرجه أعمق وفيها يشترك خبير المجال مع مهندس المعرفة حيث يسمح للخبير بإجراء مراجعات عديدة لما يريد أن يضمنه في البرنامج وإعطاء تفسيرات مطولة بنقاط معينة وإضافة معلومات جديدة وقد يتم الاستعانة بخبراء آخرين للتعليق على المعلومات التي أعطيت بواسطة الخبير الأول ويمكن توضيح حلقه التعاون بين خبير المجال ومهندس المعرفة بالشكل(٢- ٥):



الشكل (٢-٥) مرحلة تكوين المفاهيم

٣- الصياغة Formalization: يتم في هذه المرحلة حل المشكلة باستخدام أساليب الذكاء الاصطناعي، حيث يقوم مهندس المعرفة باختيار الأسلوب المناسب لتمثيل المعرفة للنظام الخبير.

٤- التطبيق Implementation: في هذه المرحلة تتم برمجة المفاهيم التي تمت صياغتها باستخدام لغات الذكاء الاصطناعي لينتج عن ذلك النموذج الأول للنظام الخبير.

٥-الاختبار Testing: يقوم مهندس المعرفة باختبار النموذج أو النظام من النواحي التالية :

أ- صحة التطبيق.

ب- كفاءة واكتمال صياغة القواعد.

ونجاح الاختبار معناه أن يقدم النظام الخبير حلولاً للمشكلات كالتي يقدمها خبيراً من البشر. لذلك لا يعتبر النظام الخبير ناجحاً إلا عندما تتطابق الحلول التي يقدمها مع حلول خبير المجال.

٢-٢-٥ خصائص النظم الخبيرة

يمكن وضع متطلبات عامة لا بد أن تتوافر في النظم الخبيرة وهي كما يلي [٢٩]:

١. أن يحتوي هيكل النظام الخبير بالدرجة الأولى على قاعدة معرفة أساسية تشتمل على شروط مماثلة للخبرة المكتسبة للإنسان وطرق معالجته لموضوع معين لكي يصل في النهاية إلى التعرف على القرار الصائب.
٢. أن يكون هذا النظام قادراً على التعامل مع قاعدة البيانات الكبيرة التي تصف الهدف.
٣. أن يشتمل النظام الخبير على أساليب بحث ذات كفاءة عالية نظراً لتعدد قواعد البيانات وقواعد المعرفة وأن يكون النظام قادراً على التفرع السريع.
٤. أن يكون النظام قادراً على التعامل مع بيانات غير كاملة وناقصة ومشوشة مثل النظم الخبيرة المستخدمة للتعرف على نوع الطائرة من بيانات توضح جزء فقط منها أو استخدام النظام لقراءة كتابة غير واضحة أو مطموسة.
٥. إمكانية إدخال بيانات لقواعد المعرفة تحتوي على شروط تمثل خبرات جديدة وذلك لتحديث النظام والوصول إلى ثقة أكبر في اتخاذ القرار وربما لبناء نظم عملاقة، ومن هنا نرى أن النظم الخبيرة هي التطور الطبيعي لنظم الذكاء الصنعي التي تحمل ذكاء وخبرة الإنسان إلى النظم والبرامج المستخدمة على الحاسبات.

٢-٢-٦ المجالات المناسبة للنظم الخبيرة

الفرق بين طبيعة المجالات المناسبة للنظم الخبيرة والتي فيها هذه النظم ضرورية بالدرجة الأولى والفرق بينهما وبين النظم الحسابية والمحاسبية تبين في الجدولين التاليين [٢٩]:

مجالات مناسبة للنظم الخبيرة	مجالات غير مناسبة للنظم الخبيرة
لا توجد حلول وقواعد خوارزمية بل توجد نظم هرمية.	توجد قواعد وحلول خوارزمية تتابعيه تعطي الحل.
يوجد عدد قليل من الخبراء في المجال.	لا يتطلب وجود خبراء في هذا المجال ويساوي ذلك وجود خبراء كثيرون من عدمه.
البيانات المتاحة مشوشة.	حقائق دقيقة ومعادلات وطرق حسابية.
مجالات تشخيصية أو في مجال التنبؤ.	تستخدم طرق الاشتقاق العددي لإعطاء الحل.

جدول (٢-١) يبين مقارنة بين مجالات النظم الخبيرة والنظم الحسابية

النظم المستخدمة في المحاسبات التقليدية	النظم المستخدمة في النظم الخبيرة
يحكم التدفق بطريقة منتظمة الخطوات.	يحكم التدفق بالبيانات والمعلومات والمعارف والرموز.
المعالجة أساساً رقمية.	المعالجة أساساً رمزية.
هيكل الحاسب يعتمد على أساس هيكل معمارية	هيكل الحاسب لا يعتمد على هيكل معمارية
التفرق المعتمد على بيانات قليلة عند المدخل لتعطي بيانات كثيرة عند المخرج.	التجمع المعتمد على بيانات كثيرة عند المدخل ليعطي نتيجة واحدة عند المخرج.
المعالجة تسير على خطوات الخوارزميات.	المعالجة تسير على الخطوات المختلفة لطرق البحث الهرمي.
يتطلب العمل على الحاسب مبرمج لوضع خطوات الخوارزميات بإحدى اللغات.	يتطلب العمل خبير معرفة لوضع الشروط واستنتاج النتائج.
معالجة تتابعية.	معالجة تحاورية ومتوازية.
الهيكل البنائي يمتد في شكل خطى كالخط المستقيم.	الهيكل البنائي تحاوري لا خطي ويمكن أن يصبح دائري الشكل.

جدول (٢-٢) يبين مقارنة بين طبيعة النظم المستخدمة في المحاسبات التقليدية والنظم المستخدمة في النظم الخبيرة

٢-٢-٧ السمات العامة للنظم الخبيرة

من أهم سمات النظم الخبيرة والتي تتميز بها عن بقية النظم الأخرى ما يلي [٢٨,٢٧] :

- تعتبر قواعد المعرفة قواعد متخصصة في موضوع أو مجال أو نطاق معين.
- تستخدم طرق الاستدلال والاشتقاق لاستخراج النتائج وهي ما تسمى مجازاً بمحرك الاستدلال.
- لاستخدام طرق الاشتقاق المتعلقة بالأسباب لا بد من تمثيل جميع القواعد والشروط على شكل مجموعات من الأدوات الشرطية (إذا) توفر الشرط (تكون) النتيجة.

مهما اختلفت النظم الخبيرة في الغرض الذي صممت من أجله فإن هناك مواصفات عامة تمنح النظام الخبير الجودة "والذكاء :

- إنه سهل الاستخدام لأي مستخدم سواء مستخدم عادي أو مطوّر.
- أن يكون نافعاً في المجال الذي صُمم له.

- أن يكون قادراً على التعليم، عندما يستخدمه غير المتخصصين، وقادراً على تطوير أداء المتخصصين ذوي الخبرة البسيطة.
- أن يكون قادراً على التعليل وشرح الأسباب للتوصيات التي يقدمها.
- أن يستجيب للأسئلة البسيطة التي يطرحها غير المتخصصين.
- أن يكون قادراً على التعلم من خلال الأسئلة التي يطرحها على المستخدم.
- أن يكون سهل التعديل، قابلاً لتحديث المعلومات أو الإضافة إليها أو إصلاح ما بها من عيوب.

٢-٢-٨ مزايا وعيوب النظم الخبيرة

تم تطبيق تكنولوجيا النظم الخبيرة خلال السنوات القليلة الماضية بنجاح في الآلاف من المنظمات عالميا على المشاكل التي امتد مداها من بحوث نقص المناعة المكتسبة إلى تحليل الغبار في المناجم. ومن أهم مزايا النظم الخبيرة:

١. زيادة الإنتاجية والمخرجات، فمثلاً في شركة DEL التي هي جزء من شركةCOMPAQ ، يقوم النظام الخبير بتخطيط المواصفات للمكونات لكل طلب للزبون، وبذلك تزيد من التهيئة للإنتاج أربعة أضعاف.
٢. زيادة الجودة، فالنظم الخبيرة تستطيع تقليل نسب الخطأ.
٣. أخذ ونشر الخبرة النادرة، فمثلاً يستخدم الأطباء في مصر والجزائر نظام خبير للعناية بالعين تم تطويره في جامعة (روتغرس) للتشخيص والتوصية بالعلاج.
٤. العمـــل في البيئة الخطرة، فالنظم الخبيرة التي تفسر المعلومات التي يتم جمعها بواسطة المستشعرات تجنب العاملين من العمل في البيئات الحارة، والرطبة أو السامة.
٥. الوصول إلى المعرفة ومكاتب المساعدة، حيث تستطيع النظم الخبيرة زيادة إنتاجية الموظفين أو حتى أتمتة عمل هذه المكاتب.
٦. زيـــادة فاعلية أنظمة أخرى، حيث يمكن أن تتكامل النظم الخبيرة مع أنظمة أخرى، مما يجعل النظم الأخرى أكثر فاعلية.
٧. القابلية على العمل مع المعلومات الناقصة أو غير المؤكدة، حتى مع الإجابة "لا اعرف" أو"غير متأكد" يستطيع النظام الخبير أن يولد إجابة على الرغم من أنها قد تكون غير مؤكدة.
٨. توفير التدريب، فالمبتدئون الذين يعملون مع النظم الخبيرة يصبحون أكثر خبرة ويعود الفضل إلى وسائل التفسير التي تخدم كوسائل للتعليم وقواعد المعرفة.
٩. تحسين اتخاذ القرار وقابلية حل المشكلة.
١٠. تقليـــل وقت اتخـــاذ القرار، فالنظم الخبيرة تستطيع اتخاذ قرارات أسرع من قرارات البشر الذين يعملون وحدهم.

١١. تقليل وقت الأعطال، حيث تستطيع النظم الخبيرة تشخيص الأعطال بسرعة وتصف عمليات التصليح. مثلاً هناك نظام خبير يسمى drilling advisor يستطيع اكتشاف الأعطال في الأجهزة النفطية حيث أنه يحقق توفيراً في كلفة وقت الأعطال بمقدار (٢٥٠٠٠٠$) في اليوم الواحد.

وبالرغم من فوائدها العديدة فإن مناهج النظم الخبيرة المتوفرة ليست دائما مباشرة ومؤثرة. وهذه بعض المشاكل التي حدّدت من الانتشار التجاري للأنظمة الخبيرة [٢٩]:

١. المعرفة التي يجب أن تكتسب ليست متوفرة دائما.
٢. الخبرة صعبة الاستخلاص من البشر.
٣. إن توجه كل خبير إلى حالة معينة قد تكون مختلفة، رغم ذلك فهي صحيحة.
٤. من الصعب حتى لخبير ذي مهارة عالية تخمين الحالات بشكل دقيق حينما يكون تحت ضغط زمني.
٥. لمستخدمي النظم الخبيرة حدود إدراك طبيعية لذلك فأنهم قد لا يستخدمون الفوائد من النظام إلى المدى الكامل.
٦. إن النظم الخبيرة تعمل جيدا فقط ضمن موضوع ضيق ومحدد مثل تشخيص العطل في آلة معينة.
٧. إن معظم النظم الخبيرة ليس لها وسائل مستقلة لفحص فيما إذا كانت الاستنتاجات معقولة وصحيحة.
٨. إن المفردات التي يستخدمها الخبراء للتعبير عن الحقائق والعلاقات تكون محددة وقد لا يفهمها الآخرون.
٩. قلة الثقة من قبل المستخدمين قد تكون عائقاً لاستخدام النظم الخبيرة.
١٠. عملية نقل المعرفة هي عرضة للتغيرات في الإدراك والحكم.
١١. إن احتمال النصيحة الخاطئة التي يوفرها النظام الخبير تكون صعبة التقييم.
١٢. بالإضافة إلى ذلك فان النظم الخبيرة قد لا تكون قادرة على الوصول إلى استنتاجات (خصوصا في المراحل الأولى من تطوير النظام).

٢-٣ التنقيب في قواعد البيانات

- إن التنقيب في قواعد البيانات يهدف إلى استخلاص المعلومات المخبأة فيها. وهي تكنولوجيا حديثة فرضت نفسها بقوة في عصر المعلوماتية وفي ظل التطور التكنولوجي الكبير وانتشار استخدام قواعد البيانات [١٥].
- استخدامها يوفر للمؤسسات في جميع المجالات القدرة على استكشاف، والتركيز على، أهم المعلومات في قواعد البيانات.
- تركز تقنيات التنقيب على بناء التنبؤات المستقبلية واستكشاف السلوك والاتجاهات، مما يسمح بتقدير القرارات الصحيحة واتخاذها في الوقت المناسب.

- تجيب تقنيات التنقيب على العديد من الأسئلة، وفي وقت قياسي، خاصةً تلك النوعية من الأسئلة التي كان من الصعب الإجابة عليها، إن لم يكن مستحيلاً، باستخدام تقنيات الإحصاء الكلاسيكية، والتي كانت إن وجدت فإنها تستغرق وقتاً طويلاً والعديد من إجراءات التحليل.

٢-٣-١ تخطيط عمليات التنقيب في قواعد البيانات

إن تخطيط عمليات التنقيب في قواعد البيانات من الأمور المهمة للحصول على أفضل النتائج، فالتخطيط الجيد يؤدي للنتائج الجيدة [١٦,١٧].

ويمكن تلخيص الخطوات الأولية للتنقيب في قواعد البيانات في ما يلي:

١ – تحديد المشكلة المراد بحثها وإيجاد الحلول لها.

٢ – بناء قاعدة بيانات التنقيب.

٣ – استكشاف البيانات.

٤ – تحضير البيانات للتنقيب.

٥ – بناء نموذج التنقيب المناسب.

٦ – تطبيق النموذج.

٧ – استخراج النتائج.

٢-٣-٢ التقنيات الحديثة للتنقيب في قواعد البيانات

١ – الجار الأقرب (Nearest Neighbor) : تعتبر من تقنيات التنقيب في البيانات التي تهدف للتنبؤ عن طريق مقارنة السجلات الشبيهة بالسجل المراد التنبؤ، وتقدير القيمة المجهولة لهذا السجل بناء على معلومات لتلك السجلات. وكثيراً ما تستخدم خوارزمية الجار الأقرب في مجال الأعمال، ومن الأمثلة الشائعة الاستخدام تلك التي تساعد المستخدمين في الشراء عن طريق اختيار السلع الأقرب لاحتياجاتهم مقارنة بسلع قد تم شراؤها بالفعل [٥٧].

٢ – التحليل العنقودي (Cluster Analysis): هي عملية تجميع السجلات المتشابهة في مجموعات، ويتم ذلك بهدف الاستكشاف العالي المستوى لما يجري داخل قاعدة البيانات. وبشكل عام، يهدف التحليل العنقودي إلى وضع العناصر المتجانسة في مجموعات منفصلة. والقاعدة العامة لضم أي عنصر في مجموعة هي أن يكون العنصر مائلاً للتشابه بعنصر منها أكثر من أن يكون شبيهاً لعنصر من مجموعة أخرى.

٣ – شجر القرار (Decision Trees): شجرة القرار هي نموذج استكشافي يظهر على شكل شجرة، كما يعبر اسمها. وبشكل دقيق يمثل كل فرع من فروعها سؤالاً تصنيفياً وتمثل أوراقها أجزاء من قاعدة البيانات تنتمي للتصنيفات التي تم بنائها. وبالرغم من أن شجرة القرار تستخدم في الاستكشاف وتحضير البيانات للعمليات الإحصائية إلا أنها أيضاً تستخدم وبشكل أكثر للتنبؤ. ومن المهم جداً عند بناء خوارزمية شجرة القرار أن يؤخذ بعين الاعتبار أن تكون قابلة للتطبيق بقدر الإمكان وبشكل مثالي على كل البيانات المتوفرة.

والقاعدة الأساسية في بناء شجرة القرار هي إيجاد أفضل سؤال عند كل فرع من فروع الشجرة بحيث يقسم هذا السؤال البيانات إلى قسمين، القسم الأول منها ينطبق عليهم السؤال والقسم الثاني لا ينطبق، وهكذا يتم من خلال سلسلة من الأسئلة بناء شجرة القرار بفروعها المتسلسلة.

٤ – الشبكات العصبية (Neural Networks): تعتبر الشبكات العصبية هي وأشجار القرار من أهم تقنيات التنقيب في البيانات، نظراً للنتائج الدقيقة التي يتم التوصل إليها باستخدام هذه الخوارزميات ولإمكانية تطبيقهما في حل العديد من المشاكل وبكافة الأنواع، هذا بالرغم من صعوبتهما والتي أدت لعدم الانتشار بشكل واسع لهما. وخوارزمية الشبكة العصبية تشبه في تركيبتها تركيبة مخ الإنسان، فهي تعمل بنفس الطريقة كما يعمل المخ في نقل ومعالجة المعلومات والتوصل إلى الاستنتاجات واكتشاف الأنماط والتنبؤات ونستطيع من خلالها تطبيق بعض ما يطبقه المخ الطبيعي، رغم أن العلماء لا يزالون حتى اليوم يكتشفون المزيد ولم يلموا بكل تفاصيل عمل مخ الإنسان.

٥ – الاستقراء (Induction): سيتم شرحه بشكل مفصل في الفقرة التالية.

٢–٤ الاستقراء Induction

عرف تعلم الآلة منذ وقت طويل كسمة أساسية للذكاء الصنعي. ويصنف ديتيرتيش وآخرون (١٩٨١) طرق التعلم في أربع مجلات: التعلم الصمي، والتعلم بالقول، والتعلم من الأمثلة، والتعلم بالتناظر [١٢].
يراجع هذا المقطع طريقة التعلم من الأمثلة لأنها تمثل الأسلوب الأكثر شيوعاً في الوقت الحاضر في تصميم النظم الخبيرة – ويدعى غالباً هذا الأسلوب بالاستقراء.

٢–٤–١ تعريف

الاستقراء هو استنتاج القواعد العامة من المعرفة المحتواة في مجموعة منتهية من الأمثلة. وبكلمات أخرى، الاستقراء هو عملية الاستنتاج المنطقي من مجموعة معطاة من الحقائق لاستنتاج المبادئ أو القواعد العامة.

على سبيل المثال، إذا أخبرت شخص ما أني أحب كرة القدم، وكرة القاعدة (البيسبول)، وكرة السلة، عندئــذ يمكن أن يستنتج ذلك الشخص بشكل صحيح بالاستقراء أني أحب الرياضة. ويبحث الاستقراء عن الأنمــاط في المعلومات المتاحة للاستدلال على نتائج منطقية معقولة.

ويكون التعلم الاستقرائي منطقة بحثية هامة في الذكاء الصنعي (AI) وفي تطــوير الــنظم الخبيــرة. وطبق وينستون (١٩٧٥) التعلم الاستقرائي للاستخلاص مفهوم الأوصاف المفهومية لهياكــل المجموعــات العالمية. واستخدم المشروع Met-Dendral أسلوب الاستقراء لاكتشاف مجموعة قواعد للاســتدلال علــى التراكيب الكيميائية من المعطيات الطيفية الكبيرة. وبين ميشالســكي وشيلوســكي (١٩٨٠) أنــه يمكــن أن يستخدم خليط من "التعلم من الأمثلة" "والتعلم بالقول" لتشكيل مجموعة من المعرفة لتشخيص أمراض فــول الصويا في النظام الخبير"PLANT" [١٢].

٢-٤-٢ الاستقراء كبحث

يمكن أن يرى التعلم الاستقرائي كبحث خلال فضاء المشكلة، أيضاً يدعى فضاء البحث، عن حل للمشــكلة. ويتألف فضاء المشكلة من المفاهيم الرئيسية للمشكلة، مرتبطة معاً بواسطة العملية الاستقرائية التي تســتخدم أمثلة المشكلة. ويشار أيضاً إلى فضاء المشكلة بشجرة القرار، والتي يمكن أن تستخدم في البحث عن حــل للمشكلة.

وبالاستقراء، يمكننا إنتاج شجرة القرار من مجموعة الأمثلة لكي تستخدم في عملية البحث، أو يمكننــا إنتاج مجموعة من قواعد القرار لكي تستخدم في النظام الخبير. ومن أجل المشاكل الكبيرة التــي تســتخدم كثير من عوامل القرار والأمثلة، يمكن أن تشكل خوارزمية الاستقراء مجموعة صغيرة من القواعــد التــي تستخلص بصورة فعالة معرفة المشكلة.

٢-٤-٣ مناهج الاستقراء

طورت عدة مناهج للاستقراء في مجال تعلم الآلة والتنقيب بالمعطيات التي تختلف في الطــرق المســتخدمة في اكتشاف المعرفة للمساعدة في مهمة اكتساب المعرفة في النظم الخبيرة. وفي هــذا المقطــع الفرعــي، نراجع بعض أساليب الاستقراء الأكثر شيوعاً.

٢-٤-٣-١ منهج ID3

طورت خوارزمية ID3 بواسطة العالم كوينلان روس (١٩٨٦). وإنها تستخدم في الوقــت الحاضــر فــي معظم أغلفة النظم الخبيرة. وتكون هذه الخوارزمية خوارزمية لبناء شجرة القرار والتــي تحــدد التصــنيف

للأشياء باختبار قيم خواصها. وإنها تبني شجرة القرار في شكل من الأعلى إلى الأسفل، ابتداء من مجموعة من الأشياء وتحديد الخواص. عند كل عقدة للشجرة، تختبر الخاصية وتستخدم النتـــائج لتجزئـــة مجموعـــة الأشياء. وتنجز هذه العملية بصورة متكررة حتى تكون المجموعة في شجرة فرعية معينة متجانسة بالنســبة لمعايير التصنيف. وبكلمات أخرى، إنها تحتوي على الأشياء المنتمية إلى نفس الفئة (الصــنف). ومــن ثــم يصبح هذا عقده ورقية. عند كل عقدة، يتم اختيار خاصية للاختبار بناء على معايير نظرية المعلومات التـــي تحاول أن تكبر كسب المعلومات وتصغر الانتروبية. وفي مصطلحات أبسط، الخاصية التي تُختبـــر تقســـم المجموعة المرشحة إلى مجموعات فرعية أكثر تجانساً [١٣,١٤]. لكن المنهج ID3 لا يعطي نتيجة دقيقـــة عندما يكون هناك الكثير من الفوضى أو العشوائية في مجموعة البيانـــات المدروســـة أو عينـــة التـــدريب وبالتالي تتم معالجة سابقة للبيانات بشكل مكثف قبل بناء شجرة القرار باستخدام النموذج ID3.

و يمكن تلخيص خوارزمية ID3 كالتالي [١٥]:

- دراسة كل الصفات وحساب الانتروبية لها اعتماداً على عينة التدريب.
- اختيار الصفة التي تملك أصغر قيمة للانتروبية.
- إنشاء عقدة عند هذه الصفة.

ID3 (Examples, Target_Attribute, Attributes)

- Create a root node for the tree
- If all examples are positive, Return the single-node tree Root, with label = +.
- If all examples are negative, Return the single-node tree Root, with label = -.
- If number of predicting attributes is empty, then Return the single node tree Root, with label = most common value of the target attribute in the examples.
- Otherwise Begin
  - A = The Attribute that best classifies examples.
  - Decision Tree attribute for Root = A.
  - For each possible value, vi, of A,
- Add a new tree branch below Root, corresponding to the test A = vi.
- Let Examples(vi), be the subset of examples that have the value vi for A
- If Examples(vi) is empty
- Then below this new branch add a leaf node with label = most common target value in the examples
- Else below this new branch add the subtree ID3 (Examples(vi), Target_Attribute, Attributes – {A})
- End
- Return Root

٢-٤-٣-٢ منهج $C_{4.5}$

اقترحت خوارزمية $c_{4.5}$ من قبل كـــوينلان (١٩٩٣). وهــي تحســين للخوارزميــة. ID3 وتولــد هــذه الخوارزمية شجرة قرار التصنيف لمجموعة معطيات معينة بالتجزئة التكرارية للمعطيــات بــنفس طريقــة خوارزمية ID3 بحيث يجري تخزين البيانات في كل عقدة من الشجرة لتحديد أفضل تقسيم للصفات. لكــن هنا يتم إجراء تحسين باستبدال العقدة الداخلية بعقدة ورقية مما يؤدي إلى تقليل نسبة الخطــأ (برودغورليــد ٢٠٠٢). وينمو القرار باستخدام إستراتيجية "العمــق – أولاً. وتأخــذ الخوارزميــة بعــين الاعتبــار كــل الاختبارات الممكنة التي يمكن أن تجزئ مجموعة المعطيات وتختار الاختيار الذي يعطــي أفضــل ربــح معلوماتي. حيث أنها تختلف عن خوارزمية ID3 بأنها تقبل كل الصفات المستمرة والمتقطعة في بناء شجرة القرار. ومن أجل كل صفة متقطعة، يؤخذ بعين الاعتبار اختيار واحد مع النتائج التي لا يتجاوز عددها القيم المختلفة للصفة. ومن أجل كل صفة مستمرة، تؤخذ بعين الاعتبار اختبارات ثنائية تستخدم كل القيم المختلفة للصفة. ولتجميع ربح الانتروبية لكل هذه الاختبارات الثنائية بصورة فعالــة، تفــرز مجموعــة معطيــات التدريب المنتمية إلى العقدة التي تحت الاعتبار من أجل قيم الصفة المستمرة ويحسب ربح انتروبية للقطــع الثنائية بناءً على كل القيم المختلفة في مسح واحد للمعطيات المفروزة. وتكرر هــذه العمليــة لكــل صــفة مستمرة [١٨,١٩].

- Check for base cases
- For each attribute a
  Find the normalized information gain from splitting on a.
- Let a_best be the attribute with the highest normalized information gain.
- Create a decision node that splits on a_best.
- Recur on the sublists obtained by splitting on a_best, and add those nodes as. children of node

٢-٤-٣-٣ منهج C 5. 0

اقترحت هذه الخوارزمية من قبل كوينلان (٢٠٠٠). وتولد هذه الخوارزميــة كــلاً مــن أشــجار القــرار ومجموعات القواعد. وتمثل C 5. 0 إعادة تفكير كامل لخوارزميات كوينلان لتوليــد مجموعــات القواعــد باستخدام C 4. 5. وتكون C 5. 0 أسرع من C 4. 5 في تشفير المعطيات، وتستخدم أقل من ١٠% مــن الذاكرة (٥. ٦ MB مقابل أكثر من٦٠ MB لــ C 4. 5)، وتنتج مجموعة قواعد أكثر دقة مع أنها تستخدم شجرة قرار أصغر بكثير. وتكبر فائدة C 5. 0 مع مجموعة المعطيات الكبيرة. وتشمل **C5.0** دعــم كامــل لأسلوب أسلوب لتوليد وتركيب المصنفات المتعددة لإعطاء دقة تنبؤية محسنة بشكل كبير، مع أي عدد مــن

التجارب. وتدمج C 5.0 عدة تسهيلات جديدة مثل تكاليف التصنيف الخاطئ للمتغيرات. في C4.5، تختبر كل الأخطاء بالتساوي. ولكن في التطبيقات العملية، تكون بعض أخطاء التصنيف أخطر من أخطاء أخرى. وتسمح C 5.0 بتعريف تكلفة منفصلة لكل زوج صفي فعلي متنبأ به. وبناء على ذلك، تبنى C5.0 المصنفات لتصغير التكاليف المتوقعة للتصنيف الخاطئ بدلاً من معدلات الأخطاء [٢٠].

٢-٤-٣-٤ منهج SLIQ

وهي (SLIQ التعلم تحت الإشراف في التحقيق)، المطورة بواسطة فريق مشروع تحقيق IBM، شجرة قرار مصممة لتصنيف معطيات التدريب الكبيرة، وإنها تستخدم أسلوب الفرز المسبق في مرحلة نمو الشجرة. وهذا يساعد على تجنب تكاليف الفرز عند كل عقدة. تحتفظ SLIQ بقائمة مفروزة منفصلة لكل صفة مستمرة وتدعى هذه القائمة، قائمة الصف، يقابل (يناظر) المدخل في قائمة الصف عنصر المعطيات ويملك علاقة الصف واسم العقدة التي ينتمي إليها في شجرة القرار. يملك المدخل في قائمة الصفات المفروزة قيمة الصفة وفهرس لعنصر المعطيات لقائمة الصف. تكبر SLIQ شجرة القرار بطريقة " العرض _ أولاً ". من أجل كل صفة، إنها يمسح القائمة المفروزة المقابلة وتحسب قيم الانتروبية لكل صفة، يتم اختيار صفة واحدة للتفريق لكل العقد في الحد الحالي وتمدد لكي تملك حد جديد. ومن ثم ينفذ أكثر من مسح واحد لقائمة الصفة المفروزة التحديد قائمة الصف لكل العقد الجديدة [٢١].

٢-٤-٣-٥ منهج SPRINT

المنهج (SPRINT خوارزمية قابلة للنمو بشكل موازي) تم تطويره من قبل شافر ١٩٩٦. وهذا المنهج يقسم عينة التدريب باستخدام تقنية العرض_أولاً إلى أن يصبح كل قسم ينتمي إلى صف أو عقدة ورقية مشابه. وتم تطويره عن المنهج SLIQ بحيث يمكن أن يُنفذ على كلا النموذجين التكراري والتسلسلي، ويختلف عنه بأنه يستخدم بنيتي معطيات الصفات التي بشكل قائمة والصفات التي بشكل رسم بياني والتي لا تستهلك ذاكرة مما يجعل من هذا المنهج مناسب لعينة تدريب كبيرة. كما أنه يعالج كل الصفات المستمرة والمتقطعة [٢٢].

٢-٤-٣-٦ منهج الاستقراء الموجه بالصفة

اكتشاف المعرفة في قواعد المعطيات أو التنقيب بالمعطيات هو بحث لاكتشاف العلاقات والنماذج العامة الموجودة لكنها غير ظاهرة في قواعد البيانات الضخمة. تم تطوير مناهج عديدة في هذا الحقل ومنها منهج الاستقراء الموجه بالصفة، الذي يقوم بتنظيم معرفة النطاق على شكل هرميات مفهوم تساعد في تعميم

مفاهيم الصفات في قواعد المعطيات العلائقية. أي أنها بشكل عام تقوم باكتشاف المعرفة في قواعد المعطيات عن طريق دمج نموذج تعلم الآلة، خصوصاً أساليب التعلم من الأمثلة، مع عمليات قاعدة المعطيات وتستخرج المعطيات المعممة من المعطيات الفعلية في قواعد المعطيات. وتلخص الفكرة الأساسية للاستقراء الموجه بالصفة بالخوارزمية التالية [٢٣]:

خوارزمية (الاستقراء الموجه بالصفة الأساسي)

المدخلات:

(١) قاعدة المعطيات العلائقية.

(٢) مهمة التعلم.

(٣) هرميات المفهوم المفضلة (اختيارية).

(٤) الشكل المفضل للتعبير عن نتائج التعلم (على سبيل المثال، عتبة التعميم).

المخرجات: القاعدة المميزة المتعلمة من قاعد المعطيات.

الطريقة: يتألف الاستقراء الموجه بالصفة الأساسي من الخطوات الأربعة التالية:

* جمع المعطيات الملائمة للمهمة
* نفذ الاستقراء الموجه بالصفة الأساسي
* بسط العلاقة المعممة
* حول العلاقة النهائية إلى قاعدة منطقية.

سنطور هذا المنهج في الفصل الثالث بدمجه مع نموذج تحليل الانحدار لتشكيل خوارزمية جديدة، تدعى استقراء نموذج تحليل الانحدار الموجه بالصفة. وستستخدم هذه الخوارزمية الجديده لاكتشاف أنماط الطقس العامة في قواعد المعطيات الطقسية العلائقية الكبيرة.

٢-٤-٣-٧ منهج المجموعات التقريبية

تكون نظرية المجموعات التقريبية، المقدمة من قبل بولاك في أوائل الثمانينات [٢٤,٢٥]، أداة رياضية جديدة للتعامل مع الغموض وعدم التأكد. وتعتمد فلسفة المجموعات التقريبية على فكرة التصنيف. سنصف هذه النظرية بالتفصيل في الفصل الرابع حيث سنستخدمها لاكتشاف القواعد من المعطيات للتنبؤ بالطقس. ولكن سنرسم فقط هنا سماتها.

برهنت منهجية المجموعات التقريبية فائدتها في كثير من التطبيقات الحقيقية. وتشمل المشاكل الرئيسية التي يمكن أن تقرب باستخدام هذه النظرية: اختزال المعطيات، أي، اختزال الصفات الزائدة،

واكتشاف اعتماديات المعطيات وتقدير معنوية المعطيات، واكتشاف التشابهات أو الاختلافات في المعطيات، واكتشاف الأنماط في المعطيات، واكتشاف علاقات ارتباط السبب – النتيجة، أي القواعد [٢٦].

يكون التطبيق الهام لهذه المنهجية، الذي يتمتع بانتباه متزايد، اكتشاف المعرفة في قواعد المعطيات (KDD) أو التنقيب بالمعطيات. وتكون إحدى المهام الأولية في هذا السياق اكتشاف وتوصيف الارتباطات الداخلية للمعطيات أو علاقات الارتباط. ويكون الوجه الآخر، الذي عُولج أيضاً باستخدام نظرية المجموعات التقريبية، اكتشاف الأنماط الشاذة أو السلوك الشاذ في المعطيات بغرض اكتشاف الخدع أو التطفل [٢٦]. ونظراً لأنه يسمح بعدم الاتساقات، ونظراً لعدم وجوب العضوية المطلقة في المجموعة، تكون الإمكانية كبيرة لمعالجة الضجيج بصورة جيدة.

٢–٥ تطوير النظام الخبير الاستقرائي

تشمل الخطوات الرئيسية في تطوير النظام الخبير الاستقرائي ما يلي:

- تحديد الهدف – يجب أن نعين أولاً هدف النظام. فسيصل البحث خلال شجرة القرار أو مجموعة القواعد إلى أحد قرارات المجموعة المنتهية. ويعين كل قرار هدف محدد مسبقاً.
- تحديد عوامل القرار – تمثل عوامل القرار عقد الصفات لشجرة القرار أو الصفات الشرطية لمجموعات القواعد. وتشمل هذه العوامل سمات المشكلة التي سنأخذها بعين الاعتبار للوصول إلى القرار النهائي. وهذا يكون غالباً المهمة الأصعب عندما نطور نظام استقرائي. وإذا كانت عوامل القرار المختارة ضعيفة أو غير كاملة، عندئذ يمكن أن تكون النتيجة النهائية غير صحيحة.
- تحديد قيم عوامل القرار – تمثل قيم عوامل القرار قيم الصفات لشجرة القرار أو القيم للصفات الشرطية لمجموعة القواعد. ويجب أن نولد قائمة القيم الممكنة لكل عامل قرار. ويكون اختيار مجموعة قيم عوامل القرار الجيدة هاماً. وإذا كانت هذه القيم غامضة أو غير كاملة، عندئذ يمكن أن يملك المستخدم مشكلة بتقديم الأجوبة. وبناءً على ذلك، يكون من الهام أن نستخدم المستخدم مبكراً في المشروع ليساعد في تحديد مجموعة القيم الفعالة.
- تحديد الحلول – يجب أن نحدد قائمة القرارات النهائية التي يمكن أن يقوم النظام بها (أي التي يمكن أن يصنفها النظام). وتكون هذه القرارات عقدة ورقية لشجرة القرار أو نتائج منطقية لمجموعة القواعد.
- تشكيل مجموعة الأمثلة (مجموعة التدريب) – تحتوي الأمثلة على معرفة المشكلة وتستخدم لصنع القرارات الذكية. وتربط هذه الأمثلة قيم عوامل القرار إلى النتائج النهائية، ويمكن أن تأتي من خبير المجال أو من سجل الأحداث الماضية. وفي اختيار مجموعة الأمثلة الأصلية، يجب أن نختار الأمثلة التي تأتي من مصدر موثق وتغطي نطاق المشكلة على نحو كامل. وإذا أتت الأمثلة من عدة

مصادر، يكون من الهام أن نحتفظ بسجل عن مصدرها. وفي حالة لم يلبي أداء النظـــام التوقعـــات، عندئذ يمكن أن يحسن حذف الأمثلة من المصدر الأقل وثوقاً الموقف.

- إنتاج شجرة القرار أو مجموعة القواعد- سنحتاج للاستخدام خوارزمية استقراء، مثل الخوارزميات الموصوفة في المقطع السابق للإنتاج شجرة القرار أو مجموعة القواعد. وفي حالة لم يظهر عامـــل قرار ما في عقد الشجرة أو مجموعة القواعد، لا تحذفه من مجموعة المعطيات. فيما بعد، وعنـــدما تضاف أمثلة أخرى، يمكن أن يظهر العامل من جديد عندما تستقرىء شجرة أو مجموعـــة قواعـــد جديدة.
- اختبار النظام- تقارن خطوة الاختبار النتائج التي تتبأ بها النظام مع دراسات الحالة الفعلية.

٢-٦ مزايا الاستقراء

- اكتشاف القواعد من الأمثلة- يمكن أن تكتشف نظم الاستقراء القواعد غير المعروفة الممكنـــة مـــن مجموعة الأمثلة. وأحياناً، أثناء تطوير النظام الخبير، سيكون خبراء النطاق قادرين علـــى وصـــف معرفة مشكلتهم. ومن ناحية ثانية قد يكونوا قادرين على تقديم (تزويد) الأمثلة الماضية عن المشكلة والتي يمكن أن يستخدموها في استقراء مجموعة قواعد صنع القرار.
- تجنب مشاكل استخراج المعرفة- يقدم الاستقراء الأسلوب الذي بواسطته يمكـــن اكتســـاب معرفـــة النظام مباشرة من خلال الأمثلة الماضية. ويمكن أن يجنب هـــذا الأســـلوب المشـــاكل المرتبطـــة بمحاولات اكتساب المعرفة مباشرة من الخبير.
- إنتاج معرفة جديدة- يمكن أن تنتج أداة الاستقراء نظام خبير لتوجيه القرارات المستقبلية، حتى مـــن خلال الخبير الذي لا يكون مدركا بشكل صريح معرفة صنع القرار. وهذا ممكنـــاً لأن الاســـتقراء يستطيع أن يكشف أنماط القرار التي قد تكون غير واضحة حتى للخبير عن المشكلة.
- اكتشاف عوامل القرار الحرجة- يكتشف الاستقراء عوامل صنع القرار الأكثر أهميـــةً. ويمكـــن أن تقود هذه الميزة إلى النظم التي تصنع القرارات على أساس بضعة عوامل فقط والتي تحسن بشـــكل كبير الذكاء وكفاءة النظام.
- إزالة عوامل القرار غير ملائمة – غالباً العوامل، التي تبدو في البداية أنها هامـــة لاتخـــاذ قـــرار نهائي، تكتشف من خلال الاستقراء أنها غير ملائمة. وهذه النتيجة يمكن أن تكون ظـــاهرة للخبيـــر وتساعد في كفاءة النظام. ويجب أن تأخذ هذه النتيجة بعين الاعتبار بحذر لأن إضافة أمثلة لاحقـــة يمكن أن تكشف عوامل قرار مستثناة سابقاً تظهر إلى السطح كاعتبارات هامة.
- اكتشاف التناقضات – نتيجة للطريقة التي تدخل فيها الأمثلة إلى نظام الاستقراء، يمكن أن تكتشـــف بعض الأغلفة بسهولة الأمثلة التي تقدم نتائج متناقضة لنفس مجموعة قيم عوامل القـــرار وترســـل تقريراً فيها إلى مطور النظام. وفي بعض الحالات، قد يكون التناقض مقبولاً. هذا يعني، بمعرفة قيم

عوامل القرار، يمكن أن تكون كل النتائج منطقية. وفي هذه الحالة، يمكن أن يخبر المطور الغلاف لأن يقبل كل الأمثلة. ولكن، في معظم الحالات، يشير التناقض المكتشف إلى مشكلة. ويمكن أن تكون المشكلة مثالاً سيئاً، حيث تكون عوامل القرار أو قيم عوامل القرار الحالية غير كافية للتفريق بين النتائج.

٢-٧ مساوئ الاستقراء

- صعوبة اختيار عوامل القرار الجيدة – تعتمد فعالية النظام على اختيار عوامل القرار الجيدة. على سبيل المثال، ما هي العوامل التي سنأخذها بعين الاعتبار عندما نطور نظام للتنبؤ بالطقس.
- صعوبة فهم القواعد – تنتج معظم أدوات الاستقراء نتائجها في شكل شجرة قرار. ومن أجل المشاكل المعقدة، قد يكون من الصعب فهم عملية القرار بالتتبع من خلال الشجرة. ويصبح هذا التتبع صعباً على وجه الخصوص عندما تستخدم الأداة لصياغة القواعد لدمجها في نظام خبير مبني على القواعد تقليدي.
- قابل للتطبيق فقط من أجل مشاكل التصنيف – يكون الاستقراء جيداً من أجل المشاكل التي فيها مجموعة قيم الصفات يمكن أن تصنف في نتيجة متوقعة معينة.

٢-٨ النظم الخبيرة الطقسية المطورة من خلال الاستقراء

طورت عدة نظم خبيرة طقسية ناجحة باستخدام أسلوب الاستقراء. ويمكن أن تستخدم هذه النظم خوارزميات استقراء مختلفة إلا أنها جميعاً تشترك في نفس السمة في أن المعرفة حول المجال تمثل بصورة أفضل في شكل الأمثلة. يقدم هذا المقطع مراجعة مختصرة لبعض من هذه النظم.

النظام WILLARD

تسبب العواصف الرعدية العنيفة في الولايات المتحدة سنوياً فقدان للأحياء وفقدان لملايين الدولارات نتيجة لتضرر الأملاك. وينجز التنبؤ بالعواصف الرعدية بواسطة علماء الأرصاد الخبيرين في مركز التنبؤ بالعواصف العنيفة الوطني (NSSFC). وتكون هذه المهمة مستهلكة للوقت وتستلزم تحليل مستمر لكميات المعطيات الضخمة.

طور نظام خبير يدعى WILLARD بناءً على أسلوب الاستقراء للمساعدة في هذه المهمة. وطور WILLARD باستخدام ١٤٠مثال للمعطيات الطقسية عن العواصف الرعدية. ويستخدم النظام هرمية مؤلفة من ٣٠ نموذجاً، وكل نموذج يكون لشجرة قرار فردية. ويسأل النظام المستخدم عن الحالات الطقسية

ذات الصلة بالمنطقة ومن ثم ينتج تنبؤ كامل مع تبرير مدعم. ويصف النظام التأكد لحدوث العاصفة الرعدية العنيفة مثل " عدم حدوث عاصفة " أو " عاصفة خفيفة " أو " عاصفة معتدلة " أو " عاصفة عالية " مع كل تنبؤ معطى بمجال احتمالي عددي.

طُوِرَ WILLARD باستخدام RULEMASTER، أداة استقراء لتوليد قواعد القرار. اختبر WILLARD خلال فترة أسبوع واحد في أواخر ربيع عام ١٩٨٤ في منطقة تشمل غرب تكساس ووسط تكساس، أوكلاهما، وكولورادو. وأثناء هذه الفترة، مرت خمس عواصف رعدية خلال المنطقة. وقورنت تنبؤات WILLARD مع التنبؤات المنجزة من قبل عالم أرصاد جوية خبير من NSSFC [٣٠].

النظام HAIL PREDICTION APPLICATION

يتنبأ هذا النظام الخبير بإنتاج البرد من عاصفة رعدية موجودة ومرصودة على السهول المرتفعة لكولورادو. ويستخدم النظام الرادار دوبلر والمنتج Z_{DR} المشتق من المعطيات كمدخلات. ويكون الخرج من النظام تصنيف للعواصف الرعدية لإنتاج البرد المتراوح من " غير معنوي " إلى " خطير ". وفي بعض الحالات، ينبه المستخدم على أرجحية الأعاصير والرياح القوية. وشكلت معرفة النظام باستخدام غلاف استقرائي. وأخيراً، طور النظام في مركز خدمة الطقس الوطني الأمريكي.

النظام ExPERT SySTEM FOR RAINFALL FORECASTING

يتنبأ هذا النظام بحدوث سقوط المطر والعمق الوسطي لسقوط المطر فوق مدينة ميلبورن وضواحيها في استراليا خلال فترة ٢٤ ساعة. وجمعت مجموعة المعطيات الطقسية من أرشيفات مكتب الكومنولث للأرصاد الجوية الاسترالي. ويستخدم النظام أساليب تعلم الآلة على مجموعات المعطيات الطقسية لاكتساب المعرفة آلياً. وأخيراً، نفذ النظام بواسطة غلاف استقرائي [١١].

الـفصـــل الثـــالـث
اكتشاف أنماط الطقس العامة من قواعد المعطيات العلائقية الكبيرة

٣-١ مقدمة

سواء من خلال المقابلة مع خبراء المجال (النطاق) أو من خلال البحث المكثف في الأدب، يكـون تحويـل المعرفة البشرية إلى نظام خبير مهمة مكثفة مضنية ومستهلكة جداً للوقت، وتعتبر غالباً عنق الزجـاج فـي تطوير النظام الخبير. ولكن، في السنوات الحديثة، ظهر مجال واسع من الأدوات المعتمدة على تقنيات تعلـم الآلة. وتنجز هذه الأدوات ما يسمى التنقيب بالمعطيات، أو اكتشاف المعرفة [٣١].

ويكون اكتشاف المعرفة عملية هامة في تطوير النظم الخبيرة. ويعرف بأنه استخراج غير تافه للمعلومـات الضمنية وغير المعروفة سابقاً ومن المحتمل مفيدة من المعطيات.

تجاوز نمو حجم وعدد قواعد المعطيات الطقسية بكثير القدرة البشرية على تحليل هـذه المطيـات الطقسية. ونتيجة لذلك، ظهرت الحاجة والفرصة لاكتشاف المعرفة من قاعد المعطيات الطقسية. ويمكـن أن يلعب مثل هذا الاكتشاف دوراً هاماً في فهم المعطيات الطقسية وفي تحديد علاقات الارتباط بين المتغيـرات [٣١]. ومؤخراً رتّب التنقيب بالمعطيات، كأحد المواضيع البحثية الواعدة جداً منذ التسعينيات من قبـل كـلاً من باحثي قواعد المعطيات والذكاء الصنعي.

في الدراسات السابقة، طورت طريقة الاستقراء الموجه بالصفة لاكتشـاف المعرفـة مـن قواعـد المعطيات بدون أخذ بعين الاعتبار تأثير العناصر الأخرى على العناصر الموجودة في قواعـد المعطيـات العلائقية هذه. وتدمج الطريقة نموذج تعلم الآلة، وخصوصاً التعلم من الأمثلة، مع عمليات قاعدة المعطيـات وتستخلص المعرفة العامة من المعطيات الفعلية في قاعدة المعطيات. ويكون المفتاح لهذا المـنهج الصـعود الموجه بالصفة لهرمية المفهوم من أجل التعميم الذي يطبق عمليات قاعدة المعطيات، ويقلـل فعليـاً درجـة التعقيد الحسابية لعملية تعلم قاعدة المعطيات [٣١].

في هذا الفصل، مددت طريقة الاستقراء الموجه بالصفة لاكتشاف المعرفـة العامـة مـن قواعـد المعطيات الطقسية العلائقية الكبيرة بدراسة تأثير عناصر الطقس على بعضها الـبعض باسـتخدام نمـوذج تحليل الانحدار. وبنيت هرميات المفهوم لعملية التعلم. ويمكن أن ينجز الاستقراء بصعود هذه الهرميات إلـى مستوى عالي للمفهوم بعد دراسة عناصر الطقس على بعضها البعض.

وينظم هذا الفصل كما يلي: قدمت بدائيات اكتشاف المعرفة في قواعـد المعطيـات العلائقيـة فـي المقطع ٣-٢. ويقدم المقطع ٣-٣ منهج نموذج تحليل الانحدار الموجه بالصفة. وأخيـراً، يختـتم الفصـل بمناقشة لمنهجنا في المقطع ٣-٤.

٣-٢ بدائيات اكتشاف المعرفة في قواعد المعطيات العلائقية

يجب أن تقدم أربعة بدائيات لتوصيف مهمة التعلم: المعطيات الملائمة للمهمة، المعرفة الخلفية، التمثـيلات المتوقعة لنتائج التعلم، وطلبات التعلم [٣٢,٣٣].

٣-٢-١ المعطيات الملائمة لعملية الاكتشاف

تخزن عادةً قاعدة المعطيات كمية كبيرة من المعطيات جزء منها يمكن أن يكون ملائماً لمهمة تعلـم. علـى سبيل المثال، لوصف سمات طقس مدينة ما، تكون فقط المعطيات المرتبطة بطقس هـذه المدينـة ملائمـة لعملية التعلم. ويمكن أن تكون المعطيات الملائمة ممتدة في عدة علاقات. ويمكـن أن يسـتخدم استفسـار لتجميع المعطيات الملائمة من قاعدة المعطيات.

ويمكن أن ترى المعطيات الملائمة للمهمة كأمثلة لعمليات التعلم. وبدون شك، يكـون الـتعلم مـن الأمثلة إستراتيجية هامة لاكتشاف المعرفة من قواعد المعطيات. وتجزئ معظم خوارزميـات الـتعلم مـن الأمثلة مجموعة الأمثلة إلى مجموعات موجبة وسالبة وتنجـز التعمـيم باسـتخدام المعطيـات الموجبـة والتخصيص باستخدام المعطيات السالبة. ولسوء الحظ، لا تخزن قاعـدة المعطيـات العلائقيـة صـراحةً المعطيات السالبة، ونتيجة لذلك، لا توجد أمثلة سالبة معينه يمكن أن تستخدم للتخصيص. وبناءً على ذلـك، تعتمد عملية استقراء قاعدة المعطيات بالدرجة الأولى على التعميم الذي يمكن أن ينجز بحذر لتجنب التعميم الزائد. ويمكن أن نكتشف أنواع قواعد كثيرة بواسطة عمليات الاستقراء مثل القواعـد المميـزة، وقواعـد التمييز الخ. تكون القاعدة المميزة تأكيد يصف مفهوم محقق من قبل كل أو أغلبيـة الأمثلـة فـي الصـف الخاضع لعملية التعلم، والذي يدعى صف الهدف. على سبيل المثال،. يمكن تلخص سمات عنصـر طقـس معين بواسطة قاعدة مميزة. تكون قاعدة التمييز تأكيد يميز مفهوم للصف المتعلم، يدعى صف الهدف، عـن الصفوف الأخرى (صفوف التباين). على سبيل المثال، لتمييز عنصر طقس واحد مـن عناصـر الطقـس الأخرى، يجب أن تلخص قاعدة التمييز السمات التي تميز عنصر الطقس هذا عن العناصر الأخـرى. فـي تعلم قاعدة مميزة، تجمع المعطيات الملائمة في صف واحد، صف الهدف، من أجل التعميم. ولتعلم قاعـدة تمييز، يكون من الضروري تجميع المعطيات في صفين، صف الهدف وصف التباين. وتقتضي المعطيـات في صفوف التباين أن المعطيات لا يمكن أن تستخدم في تمييز صف الهدف عن صـفوف التبـاين. وهـذا يعني، أنها تستخدم لاستثناء الخواص المشتركة في كل صف.

٣-٢-٢ المعرفة الخلفية

تمثل هرميات المفهوم المعرفة الخلفية الضرورية التي تتحكم في عملية التعميم. وغالباً تـنظم المسـتويات المختلفة للمفاهيم في تصنيف المفاهيم. وجزئياً يمكن أن يرتب تصنيف المفهوم وفقا للترتيب من العام إلـى الخاص. ويكون المفهوم الأعم الوصف الصفري (الابتدائي) والمفاهيم الأخص تقابل القيم الخاصة للصـفات في قاعدة المعطيات. وباستخدام هرمية المفهوم، يمكن أن تمثل القواعد المتعلمة في مصـطلحات المفـاهيم المعممة وتحدد في شكل بسيط يكون مرغوبا لمعظم المستخدمين. ويمكن أن تقدم هرميات المفهوم من قبـل مهندسي المعرفة أو من قبل خبراء النطاق. ويكون هذا معقولا حتى لقواعد المعطيات الكبيـرة لأن شـجرة المفهوم تسجل فقط قيم الصفات المنفصلة المختلفة، أو مجالات القيم الرقمية للصفة، والتي عموماً لا تكـون كبيرة جداً، ويمكن أن تدخل بواسطة خبراء النطاق. ويمكن أن تكتشف بعض هرميات المفهوم آليا وبعضها نصف آليا.

ويمكن أن تمثل كثير من المفاهيم العالية المستوى للقيم الرقمية بواسطة معطياتها التلخيصية وتخـدم كمفاهيم معممة. على سبيل المثال، قياس درجة الحرارة بين (٥ و١٠) درجات إما أن يمثل بمجـال درجـة الحرارة أو يعمم إلى (Cold). وأيضا، كثير من المفاهيم العالية المستوى للقيم المنفصلة يمكن أن تمثل فـي معطياتها التلخيصية وتخدم كمفاهيم معممة. على سبيل المثال، الأشهر الثلاثة: كانون الأول، وكانون الثاني، وشباط يمكن أن تعمم إلى فصل الشتاء. علاوة على ذلك، هرمية المفهوم للصفة يمكن أيضا أن تكتشـف أو تنقى بناء على علاقتها مع الصفات الأخرى.

٣-٢-٣ تمثيل نتائج التعلم

من وجهة النظر المنطقية، تكون كل حدودية في علاقة صيغة منطقية في شكل طبيعي وصـلي. وتوصـف علاقة المعطيات بواسطة مجموعة كبيرة من التراكيب الفصلية لهذه الأشكال الوصلية. وهكـذا، يمكـن أن تمثل كلاً من معطيات التعلم والقواعد المكتشفة إما في شكل علائقي أو في حسـاب المسـند مـن المرتبـة الأولى.

تدعى العلاقة التي تمثل نتائج تعلم مرحلية (أو نهائية) علاقة معممة مرحلية (أو نهائية). في العلاقة المعممة، تكون بعض أو كل صفاتها معطيات معممة، هذا يعني، عقد غير ورقية فـي هرميـات المفهـوم. وتطلب بعض خوارزميات التعلم من الأمثلة من القاعدة المتعلمة لأن تكون في شكل طبيعي وصلي. ويكون هذا المطلب عادةً غير معقول في قواعد المعطيات الكبيرة لأن المعطيات المعممة تحتوي غالباً على حالات مختلفة. ومن ناحية ثانية، القاعدة الحاوية على عدد كبير من التراكيب الفصلية تشير أنها في شـكل معقـد وتعميم إضافي يجب أن ينجز. وبناء على ذلك، يجب أن تمثل القاعدة المعممة النهائية إما بحدودية واحـدة (قاعدة وصلية) أو بعدد صغير من الحدوديات المقابلة لقاعدة فصلية بعدد صغير من التراكيـب الفصـلية.

ويمكن أن يسمح النظام للمستخدم لأن يحدد عتبة التعميم المفضلة، العدد الأعظمي للتراكيب الفصلية للصيغة الناتجة. على سبيل المثال، إذا قيمة العتبة حددت بثلاثة، ستتألف القاعدة المعممة النهائية من ثلاثة تراكيب فصلية على الأكثر.

يمكن التحكم بدرجة تعقيد القاعدة بواسطة عتبة التعميم. ويمكن أن تقود العتبة الكبيرة إلى قاعدة معقدة ذات تراكيب فصلية عديدة والناتج يمكن ألا يكون معمم كلياً. وتقود قيمة العتبة الصغيرة إلى قاعدة بسيطة ذات تراكيب فصلية قليلة. ومن ناحية ثانية، يمكن أن تؤدي قيم العتبة الصغيرة إلى قاعدة معممة بشكل زائد ويمكن أن تفقد بعض المعلومات القيمة. وتكون الطريقة الأفضل هي ضبط قيم العتبة ضمن مجال معقول واختيار القاعدة المعممة الأفضل بواسطة خبراء النطاق و(أو) المستخدمين.

٣-٢-٤ طلبات التعلم

لجعل عملية اكتشاف المعرفة تركز على مجموعة المعطيات المهتم بها ونستخلص المعرفة المطلوبة، يجب أن يستخدم طلب تعلم لتحفيز عملية الاكتشاف. ويجب أن يتألف طلب تعليم قاعدة المعطيات من:

(١) استفسار قاعدة معطيات يستخلص مجموعة المعطيات الملائمة.

(٢) نوع القواعد المتعلمة.

(٣) توصيف صف الهدف، ومن الممكن، صفوف التباين بناء على القواعد المتعلمة.

(٤) هرميات المفهوم المفضلة.

(٥) والشكل المفضل للتعبير عن نتائج التعلم.

لاحظ أن (٤) و(٥) يكونان اختياريان لأن هرميات المفهوم الافتراضية وعتبة التعميم الافتراضية يمكن أن يستخدمان إذا لم يحدد تفصيل صراحة.

في هذه الدراسة، سنستخدم الـــ DBMINER لكتابته طلبات التعلم.

٣-٣ منهج نموذج تحليل الانحدار الموجه بالصفة

تخزن قاعدة المعطيات الطقسية المعطيات المناخية. ولاستخلاص المعرفة العامة من قواعد المعطيات الطقسية، نحتاج عادة أن ننجز التعميم على البيانات الطقسية الفعلية. سنوضح هذا المنهج اعتمادا على عناصر درجة الحرارة، والسحب، والرطوبة، وسرعة الرياح، والضغط الجوي، وضغوط بخار الماء. تقدم المفاهيم العالية المستوى لدرجة الحرارة، والسحب، والرطوبة، وسرعة الرياح، والضغط الجوي، وضغط بخار الماء وهرمية المفهوم للفترات الفصلية في الجداول ٣-١، ٣-٢، ٣-٣، ٣-٤، ٣- ٥، ٣ -٦، والشكل ٣-١، على التوالي.

Very cold	cold	slightly cold	warm	moderate	normal	hot	very hot
<5	[5, 10)	[10, 15)	[15, 20)	[20, 25)	[25, 30)	[30, 35)	35 and up

جدول (٣-١) المفاهيم عالية المستوى لدرجة الحرارة

Clear	Fairly	Partly	Cloudy
0	(0, 2)	[2, 5)	[5, 8]

جدول (٣-٢) المفاهيم عالية المستوى للسحب

Very low	low	relatively low	normal	moderate	relatively high	high	very high
<30	[30, 40)	[40, 50)	[50, 60)	[60, 70)	[70, 80)	[80, 90)	90 and up

جدول (٣-٣) المفاهيم عالية المستوى للرطوبة

Calm	light air	light breeze	moderate breeze	strong breeze	tornado	hurricane
<1	[1, 4)	[4, 10)	[10, 22)	[22, 30)	[30, 38)	38 and up

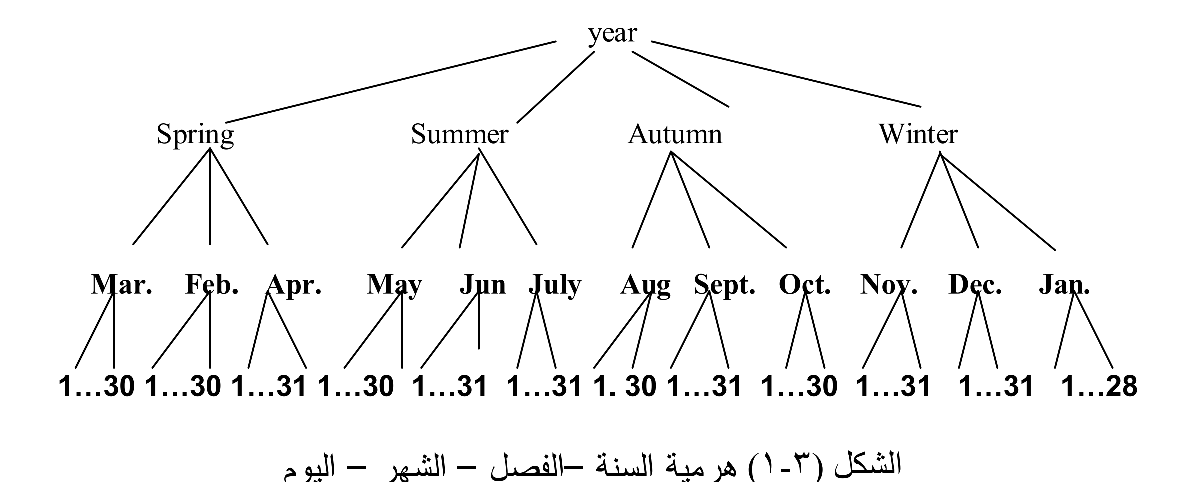
جدول (٣-٤) المفاهيم عالية المستوى لسرعة الرياح

Very low	low	relatively low	moderate	relatively high	high	very high
<1000	[1000, 1005)	[1005, 1010)	[1010, 1015)	[1015, 1020)	[1020, 1025)	1025 and up

جدول (٣-٥) المفاهيم عالية المستوى للضغط الجوي

Very low	Low	Moderate	High	Very High
<5	[5, 10)	[10, 15)	[15, 20)	20 and up

جدول (٣-٦) المفاهيم عالية المستوى لضغط بخار الماء



الشكل (٣-١) هرمية السنة –الفصل – الشهر – اليوم

نفترض أن قاعدة المعطيات الطقسيه تخزن المعطيات المناخية السابقة لمدة (١٠سنوات) من ٢٠٠١ حتى ٢٠١٠ لمحطات الطقس المبعثرة في كل أنحاء الجمهورية العربية السورية: حمص، دمشـــق، حلـــب، اللاذقية، طرطوس، حماه، دير الزور.. .. . الخ. تجمع المعطيات المناخية من هذه المحطـــات الطقســية.

وتحتوي المعطيات على المعدل اليومي لدرجات الحرارة، وكميات السحب، والرطوبة النسبية، وسرع الرياح، والضغط الجوي، وضغوط بخار الماء لكل محطة محافظة.

وتتألف عملية التعلم التي يمكن أن تقدم بالتعميم على الصفات الفعلية: درجة الحرارة، وكمية السحب، والرطوبة النسبية، وسرعة الرياح، والضغط الجوي، وضغوط بخار الماء، من الخطوات التالية:

٣-٣-١ تجميع المعطيات الطقسية الفعلية الخاصة

يستخرج تنفيذ الاستفسارات الشبيه بـــ SQL على المعطيات الطقسية الفعلية سجلات درجة الحرارة، وكمية السحب، والرطوبة النسبة، وسرعة الرياح، والضغط الجوي، وضغط بخار الماء الملائمة للمدن والأشهر والأيام. لاحظ أن " season = "winter هو جزء من معطيات معممة، الذي يجزئ إلى"month=" Dec. أو "month = "Jan. أو "month= "Feb. ، بمراجعة هرمية السنة – الفصل – الشهر –اليوم.

مهمتنا هي إيجاد نمط الطقس العام لكل يوم في فصل الشتاء اعتماداً على معطيات السنوات العشر المرصودة سابقا من ٢٠٠١ حتى ٢٠١٠. يمكن أن تبدأ عملية التعميم بالاستفسارات التالية:

In relation Table 3. 7
Learn characteristic rule for Season =“winter”
In relevance to City, Month, Day, T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9, T_{10}

In relation Table 3. 8
Learn characteristic rule for Season =“winter”
In relevance to City, Month, Day, C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9, C_{10}

In relation Table 3. 9
Learn characteristic rule for Season =“winter”
In relevance to City, Month, Day, RH_1, RH_2, RH_3, RH_4, RH_5, RH_6, RH_7, RH_8, RH_9, RH_{10}

In relation Table 3. 10
Learn characteristic rule for Season=“winter”
In relevance to City, Month, Day, WS_1, WS_2, WS_3, WS_4, WS_5, WS_6, WS_7, WS_8, WS_9, WS_{10}

In relation Table 3. 11
Learn characteristic rule for Season=“winter”
In relevance to City, Month, Day, P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}

In relation Table 3. 12
Learn characteristic rule for Season=“winter”

In relevance to City, Month, Day, VP_1, VP_2, VP_3, VP_4, VP_5, VP_6, VP_7, VP_8, VP_9, VP_{10}

حيث الجداول ٣-٧، ٣-٨، ٣- ٩، ٣ -١٠، ٣-١١، ٣ -١٢، هي جداول قاعدة المعطيات الطقسية الحاوية على المعطيات المناخية لدرجة الحرارة، وكميات السحب، والرطوبة النسبية، وسرع الرياح، والضغوط الجوية، والضغوط الجوية (بعد إضافة ١٠٠٠)، وضغوط بخار الماء لهذه المدن في فصل الشتاء، على التوالي.

City	M	D	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Homs	Dec.	1	15. 25	12. 05	19. 10	17	18. 2	18. 9	14. 7	18. 6	18. 2	15. 6
Homs	Dec.	2	16. 70	11. 80	17. 4	17. 3	16	18. 8	12. 7	16. 3	16. 5	17. 1
….												
….												
Dam.	Dec.	1	16. 15	12. 25	16. 2	16. 6	16. 5	16. 7	11. 7	14. 9	16. 2	14. 8
Dam.	Dec.	2	14. 40	12	16. 05	17. 8	13. 2	16. 1	13. 9	16. 2	16. 3	16. 4
….												
….												

جدول (٣-٧) عينة من المعطيات تمثل المعدل اليومي لدرجة الحرارة (بالدرجات) من ٢٠٠١ حتى ٢٠١٠

City	M	D	C1	C2	C3	C4	C5	C6	C7	C8	C9	C 10
Homs	Dec.	1	0. 7	3. 7	1. 5	1. 2	2. 5	1. 5	1. 2	2	0. 1	1. 6
Homs	Dec.	2	2. 4	4. 5	0. 4	2. 5	4	0. 3	3. 7	0. 5	1. 8	1. 4
….												
….												
Dam.	Dec.	1	3. 5	5. 5	4. 1	4. 2	4. 1	3. 2	5	5. 2	3. 3	3. 2
Dam.	Dec.	2	5. 7	4. 6	3. 9	4. 7	3	5. 8	4. 5	6. 9	7. 5	6
….												
….												

جدول (٣-٨) عينة من المعطيات تمثل المعدل اليومي (بالثمن) للسحب من ٢٠٠١ حتى ٢٠١٠

City	M	D	RH1	RH2	RH3	RH4	RH5	RH6	RH7	RH8	RH9	RH10
Homs	Dec.	1	50	60	55	75	77	85	63	66	45	67
Homs	Dec.	2	70	55	68	70	78	88	65	67	55	63
….												
….												
Dam.	Dec.	1	77	80	89	75	73	83	67	79	90	87
Dam.	Dec.	2	68	79	86	70	77	79	65	81	88	82
….												
….												

جدول (٣-٩) عينة من المعطيات تمثل المعدل اليومي للرطوبة النسبية (بالمائة) من ٢٠٠١ حتى ٢٠١٠

City	M	D	WS1	WS2	WS3	WS4	WS5	WS6	WS7	WS8	WS9	WS10
Homs	Dec.	1	3	4	3. 5	5	7	4. 5	2. 5	6	1. 5	11. 2
Homs	Dec.	2	3. 7	7	5. 5	4	7. 8	3. 8	8. 5	7	5	3
….												
….												
Dam.	Dec.	1	5	4. 3	6. 7	2. 5	10. 3	3. 3	4. 7	3. 9	5. 5	2. 7
Dam.	Dec.	2	6. 8	3. 9	4. 9	3	13. 1	7	8. 5	9. 5	5. 1	3. 8
….												
….												

جدول (٣-١٠) عينة من المعطيات تمثل المعدل اليومي لسرعة الرياح (بالعقدة) من ٢٠٠١ حتى ٢٠١٠

City	M	D	P1	P2	P 3	P 4	P5	P6	P7	P8	P9	P10
Homs	Dec.	1	20. 3	27. 1	19. 1	18. 4	22. 5	25	22. 5	23. 4	17. 3	18. 2
Homs	Dec.	2	24. 7	23. 1	22. 4	19. 9	20. 1	23. 6	25. 7	24. 1	18. 3	19. 2
….												
….												
Dam.	Dec.	1	25. 1	22. 3	22. 6	22. 3	19. 4	18. 5	21. 9	22. 4	14. 8	20. 3
Dam.	Dec.	2	21. 4	19. 3	20. 4	21. 1	16. 5	24. 1	27. 5	17. 3	22. 4	18. 2
….												
….												

جدول (٣-١١) عينة من المعطيات تمثل المعدل اليومي للضغط الجوي(بالميلي بار)من ٢٠٠١ حتى ٢٠١٠

City	M	D	VP1	VP2	VP 3	VP 4	VP5	VP6	VP7	VP8	VP9	VP10
Homs	Dec.	1	20	11	12	8	10	15	12	14	17	18
Homs	Dec.	2	18	13	10	9	14	16	15	11	13	19
….												
….												
Dam.	Dec.	1	21	19	16	13	17	18	15	14	18	20
Dam.	Dec.	2	18	17	18	15	19	20	17	13	19	18
….												
….												

جدول (٣-١٢) عينة من المعطيات تمثل المعدل اليومي لضغط بخار الماء(بالعقدة) من ٢٠٠١ حتى ٢٠١٠

٣-٣-٢ تنفيذ نموذج تحليل الانحدار

ننفذ تحليل الانحدار البسيط أو المتعدد على البيانات الطقسية الفعلية المجمعة لحساب القيم التقديرية لدرجة الحرارة، وكمية السحب، والرطوبة النسبية، وسرعة الرياح، والضغط الجوي، وضغط بخار الماء.

٣-٣-٢-١ نموذج تحليل الانحدار

نموذج تحليل الانحدار هو المعادلة الرياضية التي تصف العلاقة بين متغيرين أو أكثر. يدعى دراسة تأثير متغيرين مستقلين أو أكثر على متغير تابع بالانحدار المتعدد. ومن ناحية ثانية، إذا اخترنا متغير مستقل واحد فقط ودرسنا تأثير هذا المتغير علي المتغير التابع إن ذلك يدعى انحدار بسيط [٣٤].

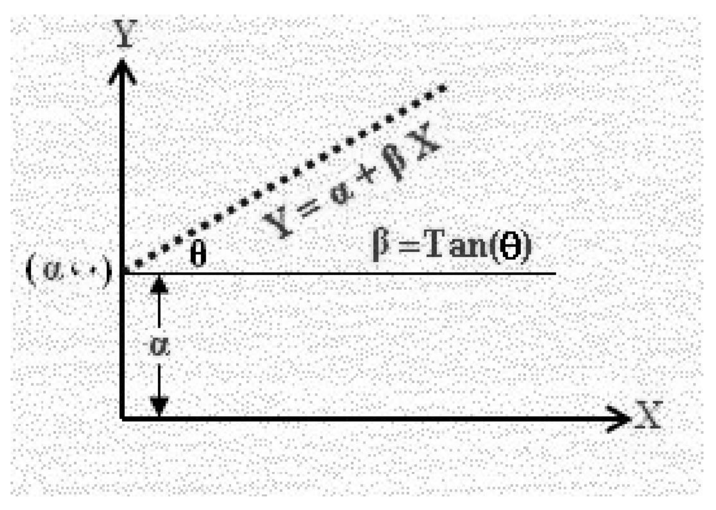
أولاً- الانحدار الخطي البسيط

تهدف دراسة الانحدار التنبؤ بقيمة متغير (y) بمعرفة متغير آخر (x) ويعرف المتغير الأول بالمتغير التابع (dependent) ويرمز له y ويقاس دون خطأ في حين يعرف المتغير الآخر بالمتغير المستقل (Independent) ويرمز له x [٣٥].

مثال: نريد أن نتنبأ بقيمة درجة الحرارة (y) عند معرفة قيمة الضغط الجوي (x) أي نريد عملياً أن نعرف العلاقة التي تربط بينهما. فإذا أعطينا قيمة ما (أي قيمة تنتمي لمجموعة الأعداد الحقيقية) للمتغير x في المعادلة $y = \alpha + \beta x$ فنحصل علي قيمة مناظرة للمتغير y فهنا قيمة y تتحدد بمعرفة قيمة x فلذا المتغير x عرف بالمتغير المستقل في حين y تتعين قيمتها تبعاً لقيمة x لذا عرفت y بالمتغير التابع (أي تبعاً لقيمة x).

كما أن الانحدار هنا بسيط لوجود متغيرين فقط تابع ومستقل، وسنتحدث لاحقاً عن الانحدار المتعدد بوجود متغير تابع واحد فقط مع وجود متغيرين مستقلين أو أكثر، وعند ذكر كلمة الخط نعني بها خط الانحدار. والانحدار يعني البحث عن هذه المعادلة أو العلاقة بين المتغيرين x المستقل، y التابع أو المعتمد كما أن المعادلة $y = \alpha + \beta x$ تحوي α, β وهما قيمتان ثابتتان حيث β تبين ميل الخط المستقيم $y = \alpha + \beta x$ (الزاوية التي يصنعها المستقيم مع الاتجاه الموجب لمحور السينات) كما في الشكل (٣-٢) وإن إشارة α تعني:

- أن تكون β موجبة فإن علاقة y بالمتغير المستقل x علاقة طردية موجبة (تزداد قيم y بزيادة قيم x المناظرة لها أو العكس أي تنقص بنقصانها).
- أن تكون β موجبة فإن الخط $y = \alpha + \beta x$ يصنع زاوية حادة مع محور السينات الموجب كما مبين بالشكل.
- أن تكونβ = صفراً فتنعدم العلاقة الخطيبة (لا توجد علاقة) وأن قيمة y ثابتة ($y = \alpha$).
- أن تكون $\beta = \infty$ فتنعدم العلاقة الخطيبة (لا توجد علاقة) كما في الشكل.
- تعرف β بميل الانحدار.



الشكل (٣-٢) ميل الانحدار الخطي البسيط

- أن تكون β سالبة فإن العلاقة عكسية سالبة (تزداد قيم y بنقص قيم x المناظرة لها أو العكس).
- في حين أن α تبين قيمة الجزء المقطوع من محور الصادات (الرأسي) بالمستقيم $y = \alpha + \beta x$
- الخط $y = \alpha + \beta x$ يمر بالنقطة (٠، α) أو أن الخط $y = \alpha + \beta x$ يمر بالزوج (٠، α)
- تعرف α بثابت الانحدار.
- عند استخدام عينه n من الأزواج من مجتمع ذو بعدين فنكتب العلاقة بحروف صغيرة $y = a + b x$ ويكون a تقديراً لـــ α وb تقديراً لـــ β.

وتسمى المعادلة $y = a + b x$ بمعادلة انحدار y على x في حين $x = a + b y$ معادلة x على y سواء للعينة أو المجتمع. ولمعرفة المعادلة يجب معرفة كل من a, b ويجدر هنا القول إن اهتمامنا يكون على العلاقة نفسها ببين x, y وليس على سبب وجود العلاقة أو الظروف المحيطة بها فهناك علاقة سببية بين كمية السماد وكمية الناتج الزراعي للقمح مثلاً. ولكننا لا يمكن أن نجزم بوجود علاقة سببية بين عدد السيارات وكمية الناتج الزراعي.

معادلة الانحدار البسيط

كما ذكرنا سابقاً، في معادلة الانحدار الخطي البسيط، يكون المتغير التابع مرتبطا بمتغيــر مســتقل واحــد. ويمكن أن تكتب معادلة الانحدار الخطي البسيط بالصورة التالية:

$$\hat{y} = a + bx \qquad (١)$$

حيث $\hat{y}$ هو التقدير المحسوب للمتغير التابع.

x هي قيمة معطاة للمتغير المستقل.

a هو قيمة $\hat{y}$ عندما يكون x تساوي الصفر.

b هو ميل خط الانحدار.

لحساب a و b سنستخدم طريقة المربعات الصغرى التي تقول: ابحث عن قيمة a و b التي تجعل المقدار $\sum(y-\hat{y})^2$ أصغر ما يمكن. يكون $\sum(y-\hat{y})^2$ مجموع الأخطاء المر بعه لكل قيمة من خط التقـــدير ويرمز له بالرمز SSE، أي:

$$SSE= \sum(y-\hat{y})^2$$
$$SSE= \sum[y-(a+bx)]^2$$

يمكن اعتبار SSE كدالة لمتغيرين a و b. لتصغير SSE سنجد $\frac{\partial SSE}{\partial a}$ و $\frac{\partial SSE}{\partial b}$ ونضع هذه المشـــتقات الجزئية مساوية إلى الصفر.

$$SSE= \sum\left[y^2-2y(a+bx)+(a+bx)^2\right]$$
$$SSE= \sum y^2-2\sum y(a+bx)+\sum (a+bx)^2$$
$$SSE= \sum y^2-2a\sum y-2b\sum xy+\sum\left(a^2+2abx+b^2x^2\right)$$

وهكذا:

$$SSE= \sum y^2-2a\sum y-2b\sum xy+na^2+2ab\sum x+b^2\sum x^2$$

حيث n من ١ إلى ١٠.

بالمفاضلة بالنسبة إلى a,b:

$$\frac{\partial SSE}{\partial a}= -2\sum y+2an+2b\sum x$$
$$\frac{\partial SSE}{\partial b}=-2\sum xy+2a\sum x+2b\sum x^2$$

بوضع هذه المعادلات مساوية إلى الصفر نحصل على المعادلات الطبيعية:

$$\sum y=an+b\sum x \qquad (٢)$$
$$\sum xy=a\sum x+b\sum x^2 \qquad (٣)$$

يمكننا اختزال المعادلتين (٢) و(٣) إلى معادلة واحدة بأخذ الانحرافات لـِ y و x عن متوســطاتها [٣٥]. في هذه الحالة، تصبح المعادلتان الطبيعيتان معادلة واحدة فقط (حيث المعادلة الأولى تصبح مســاوية إلــى الصفر) والتي يمكن أن تُحل في المجهول b. ومن ثم نعوض b في معادلة المتوسط، أي في المعادلة:

$$\bar{y}=a+b\bar{x} \qquad (٤)$$

نحصل على a.

بتعويض y و x بواسطة انحرافاتها عن المتوسط، نحصل على المعادلة التالية:

$$\sum\left(x-\bar{x}\right)\left(y-\bar{y}\right)=b\sum\left(x-\bar{x}\right)^{2} \qquad (٥)$$

من (٤) نحصل على:

$$b=\sum\left(y-\bar{y}\right)\left(x-\bar{x}\right)/\sum\left(x-\bar{x}\right) \qquad (٦)$$

من (٣) لدينا:

$$a=\bar{y}-b\bar{x} \qquad (٧)$$

بتعويض (٦) في (٧) نحصل على a. وبتعويض a و b في (١) نحصل على معادلة الانحدار البسيط.

ثانياً – الانحدار المتعدد

سبق أن تكلمنا عن الانحدار البسيط ممثلاً في معادلة خط انحداره $y = \alpha + \beta x$ والتي تظهر المتغير التابع y معتمداً على المتغير المستقل x وسندرس الآن المتغير التابع مرتبطاً بأكثر من متغير مستقل وهو ما يمكن معالجته بالانحدار البسيط بتأثير كل المتغيرات المستقلة معاً على المتغير التابع وهو ما يقودنا للانحدار المتعدد (أكثر من متغير مستقل $x1, x2,... , xn$) بقصد التنبؤ بالمتغير التابع والذي يأخذ معادلة انحداره الشكل التالي:

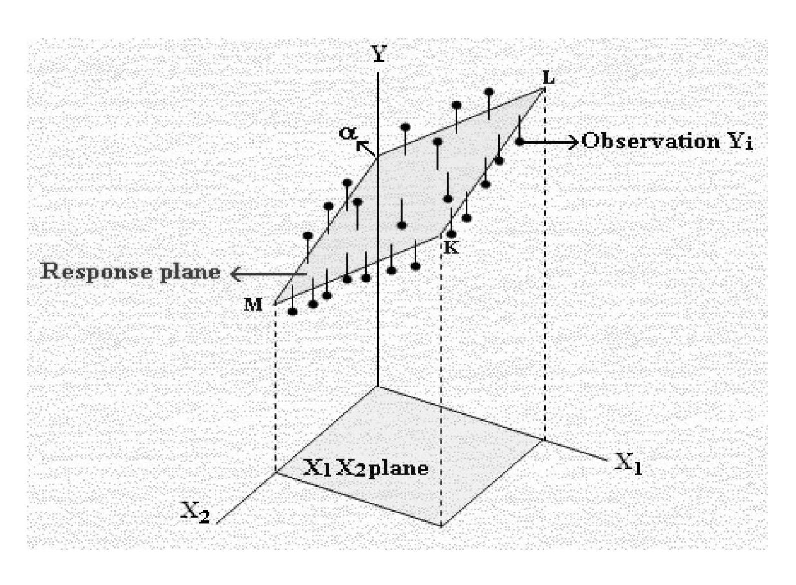
$$y = \alpha + \beta 1x1+ \beta 2x2 + \beta 3x3+..\ . + \beta kxk$$

مثال: نريد أن نتنبأ بدرجة الحرارة (y) حسب قيمة الضغط الجوي وضغط بخار الماء وسرعة الرياح وحالة الغيوم ودرجة الرطوبة أي نريد أن نعرف العلاقة التي تربط بين هذا العنصر وباقي العناصر.

وإذا أخذنا المعادلة:

$$y = \alpha + \beta 1x1+ \beta 2x2$$

حيث $x1, x2$ متغيران مستقلان، y المتغير التابع، $\alpha, \beta 1, \beta 2$ ثوابت وأن $\beta 1$ معامل انحدار $x1/ y$ وأن $\beta 2$ معامل انحدار $x2 / y$ ومثلناها في الفراغ ثلاثي الأبعاد كما مبين بالشكل (٣-٣)



الشكل (٣-٣) تمثيل الانحدار المتعدد

معادلة الانحدار المتعدد

كما ذكرنا سابقاً، في معادلة الانحراف الخطي المتعدد، يكون المتغير التابع مرتبطاً بأكثر من متغير واحد مستقل. معادلة الانحدار الخطي المتعدد هي [٣٦]

$$y^{\wedge} = a + b_1x_1 + b_2x_2 + ...+ b_{\kappa}x_{\kappa} \qquad (٨)$$

حيث $y^{\wedge}$ هي القيمة التقديرية للمتغير التابع.

x_i هي قيمة المتغيرات المستقلة.

a هي تقاطع $-y$.

b_i هي الميل المرتبط لـــِ x_i.

لتوضيح كيف نحسب a وb_i، نأخذ حالة خاصة من أجل k=2 نحصل على:

$$y^{\wedge} = a + b_1x_1 + b_2x_2 \qquad (٩)$$

لتقدير a وb_1 وb_2 نحتاج إلى ثلاث معادلات طبيعية. بصورة مشابه إلى الانحدار البسيط، يمكن أن نجد المعادلات التالية [٣٦]:

$$\sum y = na + b_1\sum x_1 + b_2\sum x_2 \qquad (١٠)$$

$$\sum x_1y = a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2 \qquad (١١)$$

$$\sum x_2y = a\sum x_2 + b_1\sum x_1x_2 + b_2\sum x_2^2 \qquad (١٢)$$

يمكننا أن نختصر المعادلات (١٠) و(١١) و(١٢) إلى معادلتين بأخذ الانحرافات لـــِ y وx_2 وx_1 عن متوسطاتها [٣٥]. في هذه الحالة، تصبح المعادلات الطبيعية الثلاث معادلتين فقط (حيث المعادلة الأولى تصبح مساوية للصفر) اللتان يمكن أن تحلان بالمجهولين b_1 وb_2. ومن ثم نعوض عن b_1 وb_2 في معادلة المتوسط، أي في المعادلة:

$$y^{-} = a + b_1x_1^{-} + b_2x_2^{-} \qquad (١٣)$$

نحصل على a.

بتعويض عن y وx_1وx_2 بواسطة انحرافاتها عن المتوسطات، نحصل على المعادلات التالية:

$$\sum(x_1 - x_2)(y - y^{-}) = b_1\sum(x_1 - x_1^{-})^2 + b_2\sum(x_1 - x_1^{-})(x_2 - x_2^{-}) \qquad (١٤)$$

$$\sum(x_2 - x_2^{-})(y - y^{-}) = b_2\sum(x_2 - x_2^{-})^2 + b_1\sum(x_1 - x_1^{-})(x_2 - x_2^{-}) \qquad (١٥)$$

بحل (١٤) و(١٥) نحصل على b_1, b_2 بتعويض b_1, b_2 في (١٣)، نحصل على a.

بتعويض a وb_1 في (٩) نحصل على معادلة الانحراف المتعدد.

٣-٣-٢-٢ اختبار المعنوية لنموذج تحليل الانحدار

لإنجاز اختبار المعنوية لنموذج تحليل الانحدار، نتبع الخطوات التالية [٣٥]:

١. احسب $SSY = \sum (y - \bar{y})^2$ –المجموع الكلي للمربعات.

٢. احسب $SSE = \sum (y - y^\wedge)^2$ – مجموع البواقي للمربعات.

٣. احسب $MS\,Regression = (SSY - SSE)/K$ – متوسط مجموع الانحدار للمربعات.

٤. احسب $MS\,Residual = SSE/n - k - 1$ – متوسط مجموع البواقي للمربعات.

٥. احسب $F - Test = \frac{MS\,Regression}{MRresidual}$

٦. قارن قيمة F المحسوبة مع القيمة الجد ولية لـِ $F_{\%95}(k, n-k-1)$. وإذا كانت $F > F_{\%95}(k, n-k-1)$، عندئذ يكون تحليل الانحدار معنويا. وبناء على ذلك، تكون القيمة المقدرة لـِ y، أيضا تدعى القيمة المتنبأ بها، هي $y^\wedge$ من أجل أي قيمة لـِ x_i.

٣-٣-٣ إنجاز التعميم الموجه بالصفة

يعمم إنجاز التعميم الموجة بالصفة على معطيات الطقسية الفعلية المجمعة قيم صفة درجة الحرارة، وقيم صفة السحب، وقيم صفة الرطوبة النسبية، وقيم صفة سرعة الرياح، وقيم صفة ضغط بخار الماء، بتقدير قيمة درجة الحرارة وقيمة كمية السحب وقيمة الرطوبة النسبة وقيمة سرعة الرياح وقيمة الضغط الجوي وقيمة ضغط بخار الماء لكل يوم في الشهر. يوضع جزء من الجداول مع القيم المقدرة والمفاهيم العالية المستوى في الجداول ٣-١٣، ٣-١٤، ٣-١٥، ٣-١٦، ٣-١٧، ٣-١٨.

City	M	D	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	$\hat{T}$	High-level concepts
Homs	Dec.	1	15. 2	12. 5	19. 1	17	18. 2	18. 9	14. 7	18. 6	18. 2	15. 6	14. 7	Slightly cold
Homs	Dec.	2	16. 7	11. 8	17. 4	17. 3	16	18. 8	12. 7	16. 3	16. 5	17. 1	17. 1	Warm
….														
….														
Dam.	Dec.	1	16. 1	12. 2	16. 2	16. 6	16. 5	16. 7	11. 7	14. 9	16. 2	14. 8	13. 5	Slightly cold
Dam.	Dec.	2	14. 4	12	16. 0	17. 8	13. 2	16. 1	13. 9	16. 2	16. 3	16. 4	14. 3	Slightly cold
….														
….														

جدول (٣-١٣) القيم المقدرة والمفاهيم العالية المستوى لدرجة الحرارة

City	M	D	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	$\hat{C}$	High-level concepts
Homs	Dec.	1	0. 7	3. 7	1. 5	1. 2	2. 5	1. 5	1. 2	2	0. 1	1. 6	1. 82	Fairly
Homs	Dec.	2	2. 4	4. 5	0. 4	2. 5	4	0. 3	3. 7	0. 5	1. 8	1. 4	2. 3	Partly
….														
….														
Dam.	Dec.	1	3. 5	5. 5	4. 1	4. 2	4. 1	3. 2	5	5. 2	3. 3	3. 2	3. 1	Partly
Dam.	Dec.	2	5. 7	4. 6	3. 9	4. 7	3	5. 8	4. 5	6. 9	7. 5	6	5. 2	Cloudy
….														
….														

جدول (٣-١٤) القيم المقدرة والمفاهيم العالية المستوى للسحب

City	M	D	RH_1	RH_2	RH_3	RH_4	RH_5	RH_6	RH_7	RH_8	RH_9	RH_{10}	$\hat{RH}$	High-level concepts
Homs	Dec.	1	50	60	55	75	77	85	63	66	45	67	65	moderate
Homs	Dec.	2	70	55	68	70	78	88	65	67	55	63	61	moderate
….														
….														
Dam.	Dec.	1	77	80	89	75	73	83	67	79	90	87	75	relatively high
Dam.	Dec.	2	68	79	86	70	77	79	65	81	88	82	71	relatively high
….														
….														

جدول (٣-١٥) القيم المقدرة والمفاهيم العالية المستوى للرطوبة النسبية

City	M	D	WS_1	WS_2	WS_3	WS_4	WS_5	WS_6	WS_7	WS_8	WS_9	WS_{10}	$\hat{WS}$	High-level concepts
Homs	Dec.	1	3	4	3. 5	5	7	4. 5	2. 5	6	1. 5	11. 2	4. 78	Light breeze
Homs	Dec.	2	3. 7	7	5. 5	4	7. 8	3. 8	8. 5	7	5	3	6. 2	Light breeze
….														
….														
Dam.	Dec.	1	5	4. 3	6. 7	2. 5	10. 3	3. 3	4. 7	3. 9	5. 5	2. 7	5. 5	Light breeze
Dam.	Dec.	2	6. 8	3. 9	4. 9	3	13. 1	7	8. 5	9. 5	5. 1	3. 8	6. 1	Light breeze
….														
….														

جدول (٣-١٦) القيم المقدرة والمفاهيم العالية المستوى لسرعة الرياح

City	M	D	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	$\hat{P}$	High-level concepts
Homs	Dec.	1	20. 3	27. 1	19. 1	18. 4	22. 5	25	22. 5	23. 4	17. 3	18. 2	23. 3	High
Homs	Dec.	2	24. 7	23. 1	22. 4	19. 9	20. 1	23. 6	25. 7	24. 1	18. 3	19. 2	20. 1	High
….														
….														
Dam.	Dec.	1	25. 1	22. 3	22. 6	22. 3	19. 4	18. 5	21. 9	22. 4	14. 8	20. 3	21. 9	High
Dam.	Dec.	2	21. 4	19. 3	20. 4	21. 1	16. 5	24. 1	27. 5	17. 3	22. 4	18. 2	21. 8	High
….														
….														

جدول (٣-١٧) القيم المقدرة والمفاهيم العالية المستوى للضغط الجوي

City	M	D	VP1	VP2	VP 3	VP 4	VP5	VP6	VP7	VP8	VP9	VP10	$\hat{VP}$	High-level concepts
Homs	Dec.	1	20	11	12	8	10	15	12	14	17	18	12. 5	Moderate
Homs	Dec.	2	18	13	10	9	14	16	15	11	13	19	15	High
….														
….														
Dam.	Dec.	1	21	19	16	13	17	18	15	14	18	20	16. 1	High
Dam.	Dec.	2	18	17	18	15	19	20	17	13	19	18	17. 3	high
….														
….														

جدول (٣-١٨) القيم المقدرة والمفاهيم العالية المستوى لضغط بخار الماء

نظراً لان قيمة درجة الحرارة المتوسطة (المقدرة) في الجدول ٣-١٣ تحتوي على قيم مختلفة عديدة ولاتصل إلى مستوى المفهوم المطلوب(الواحد)، نحتاج لأن ننجز التعميم على معطيات درجة الحرارة الفعلية. وفي هذه الحالة، تعمم قيمة درجة الحرارة المقدرة إلى مستوى أعلى مثل "slightly cold" أو"warm" الخ. بمراجعة هرمية المفهوم لدرجة الحرارة.

نظراً لان قيم السحب المتوسطة في الجدول ٣-١٤ تحتوي على قيم مختلفة عديدة ولا تصل إلى مستوى المفهوم المطلوب، نحتاج لأن ننجز التعميم على معطيات السحب الفعلية. في هذه الحالة، تعمم قيمة السحب المقدرة إلى مستوى أعلى مثل "Fairly" أو" Partly " الخ. بمراجعة هرمية المفهوم للسحب.

نظراً لان قيم الرطوبة النسبية المتوسطة في الجدول ٣-١٥تحتوي على قيم مختلفة عديدة ولاتصل إلى مستوى المفهوم المطلوب، نحتاج لأن ننجز التعميم على معطيات الرطوبة النسبية الفعلية. في هذه الحالة، تعمم قيمة الرطوبة النسبية المقدرة إلى مستوى أعلى مثل "moderate " أو "relatively high" الخ. بمراجعة هرمية المفهوم للرطوبة النسبية.

نظراً لان قيم سرعة الرياح المتوسطة في الجدول ٣-١٦ تحتوي على قيم مختلفة عديدة ولا تصل إلى مستوى المفهوم المطلوب، نحتاج لأن ننجز التعميم على معطيات سرعة الرياح الفعلية. في هذه الحالة، تعمم قيمة سرعة الرياح المقدرة إلى مستوى أعلى مثل "light breeze" الخ. بمراجعة هرمية المفهوم لسرعة الرياح.

نظراً لان قيم الضغط الجوي المتوسط في الجدول ٣-١٧ تحتوي على قيم مختلفة عديدة ولاتصل إلى المفهوم المطلوب، نحتاج لأن ننجز التعميم على معطيات الضغط الجوي الفعلية. في هذه الحالة، تعمم قيمة الضغط الجوي المقدرة إلى مستوى أعلى مثل "high" الخ. بمراجعة هرمية المفهوم للضغط الجوي.

نظراً لان قيم ضغط بخار الماء في الجدول ٣-١٨ تحتوي على قيم مختلف عديدة ولاتصل إلى مستوى المفهوم المطلوب، نحتاج لأن ننجز التعميم على معطيات ضغط بخار الماء الفعلية. في هذه الحالة، تعمم قيمة ضغط بخار الماء المقدرة إلى مستوى أعلى مثل "moderate" أو "high" الخ. بمراجعة هرمية المفهوم لضغط بخار الماء.

٣-٣-٤ دراسة الاعتقاد في التعميم

تعرض المعرفة المعممة في التعميم كمجموعات ضبابية. في المنطق الثنائي القيمة، تكون القيمة "صحيح" والقيمة "خطأ" الإمكانيات فقط. والشيء إما أن يكون في المجموعة أولا يكون[٣١]. تقاس درجة العضوية في المجموعة الضبابية A بدالة العضوية التي تعرف بالصورة التالية:

$$\mu_A(x): X \to [0,1]$$

تطبق دالة العضوية هذه x إلى نطاق من الأعداد الحقيقية المعرفة في الفترة من (٠) إلى (١) ويرمز له بالرمز[0,1]. هذا يعني، تكون دالة العضوية عدد حقيقي$0 \le \mu_A(x) \le 1$ حيث (٠) يعني عدم العضوية و(١) يعني العضوية الكاملة في المجموعة. تدعى القيمة العملية لدالة العضوية، مثل 5 .0، بعلامة (درجة) العضوية. لحساب درجة العضوية، التي تمثل الدرجة التي بها ينتمي العنصر إلى المجموعة الضبابية، سنستخدم الدالة المثلثية التالية[٣٧]:

$$\mu(x,a,m,b) = \max\{\min[(x-a)/(m-a),(b-x)/(b-m)],0\}$$

حيثm هي قيمة مساعدة وa,b يرمزان إلى الحدين السفلي والعلوي، على التوالي. في حالتنا، تكون x القيمة المقدرة التي تأتي إما من معادلة الانحدار البسيط أو المتعدد. سنفسر علامات العضوية للمعرفة العامة كأنها عوامل ثقة (CFS) التي لا تعكس فقط انحرافاتها عن الطبيعي ولكن أيضا كيف تكون معلوماتنا المؤكدة حولها [٣١]، أي: $CF = \mu_A(x)$.

عموما، يمكن أن تلخص عملية التعلم الموصوفة أعلاه في خوارزمية نموذج تحليل الانحدار الموجه بالصفة التالية [٣٨]:

خوارزمية (استقراء نموذج تحليل الانحدار الموجه بالصفة)

الدخل:

١. قاعدة معطيات علاقية
٢. مهمة التعلم
٣. مجموعة هرميات المفهوم

٤. الشكل المفضل للتعبير عن نتائج التعلم (على سبيل المثال، حد التعميم).

الخرج:

القاعدة المميزة المتعلمة من قاعدة المعطيات.

الطريقة:

١. جمع مجموعة المعطيات الملائمة للمهمة باستفسار SQL.

٢. أنجز نموذج تحليل الانحدار لدراسة تأثير العناصر على بعضها البعض.

٣. أنجز الاستقراء الموجة بالصفة على نحو متكرر حتى يكون عدد الحدوديات ضمن عتبة التعميم أو حتى تعمم كل صفة إلى مستوى المفهوم المطلوب.

٤. أخرج القاعدة المعممة.

٥. ادرس الاعتقاد في التعميم.

٣-٤ ملاحظات ختامية

لقد درسنا في هذا الفصل متطلبات وضرورة اكتشاف أنماط الطقس العامة في قواعد المعطيات الطقسية العلائقية الكبيرة وطورنا تقنية اكتشاف معرفة: استقراء نموذج تحليل الانحدار الموجة بالصفة.

تبين دراستنا أن اكتشاف أنماط الطقس العامة يمكن أن ينجز بصورة فعالة في قواعد المعطيات الطقسية العلائقية بدمج أسلوب اكتشاف المعرفة في قواعد المعطيات العلائقية، الاستقراء الموجه بالصفة، مع نموذج تحليل الانحدار لدراسة تأثير عناصر الطقس المختلفة على بعضها البعض. وينجز التعميم على المعطيات الطقسية بصعود هرمية المفهوم إلى مستوى مفهوم عالي بعد دراسة تأثير عناصر الطقس المختلفة على بعضها البعض. وأيضا يقدم هذا المنهج تقديرات للاعتقاد في التعميم باستخدام المنطق الضبابي.

الفصل الرابع

اكتشاف القواعد من المعطيات للتنبؤ بالطقس باستخدام نظرية المجموعات التقريبية (تصميم و تنفيذ و تجريب الخوارزميات)

٤-١ مقدمة

يستخدم الخبراء البشريون التجريبيات أو قواعد الإبهام للقيام بالتنبؤات الطقسية. وتكون هذه التجريبيات معتمده على عدد من العوامل الطقسية. وعينة من هذه التجريبيات تكون كما يلي: إذا كانت درجة الحرارة باردة جداً والسماء غائمة واتجاه الرياح شمالي وسرعة الرياح هادئة والضغط الجوي عالي، عندئذ يكون الطقس ماطراً.

لكي نطور نظام خبير للتنبؤ بالطقس يجب أن نكتسب المعرفة من خلال المقابلات المهيكلة وغير المهيكلة مع الخبراء البشريين ونحصل على التجريبيات لعمليات التنبؤ بالطقس. ولكن، حتى الخبراء الذين يتنبئون بالطقس اعتماداً على خبرتهم، يفهمون حاليا بشكل ضعيف عملية التنبؤ بالطقس. يتحول الافتقار إلى المعرفة عن التنبؤ بالطقس إلى فجوة في قاعدة المعرفة للنظام الخبير. بكلمات أخرى، تكون عملية اكتساب المعرفة اليدوية نفسها غير ملائمة (غير كافية) لمعالجة كل المواقف التي تظهر في تطبيق معقد للتنبؤ بالطقس.

تكون الطريقة البديلة لاكتساب المعرفة التعلم المؤتمت من المعطيات المرصودة (المشاهدة). هذا يعني، تصميم الخوارزمية التي تستطيع أن تعلم وتنقي قواعد القرار من مجموعة من عينات التدريب، أيضا تدعى مجموعة الأمثلة أو المعطيات المرصودة. ويشار إلى هذه الطريقة، كما ذكر سابقاً، بالتعلم الاستقرائي أو التعلم من الأمثلة.

في هذا الفصل، نقدم تطبيق نظرية المجموعات التقريبية للاكتشاف الآلي للقواعد من مجموعة من عينات المعطيات للتنبؤ بالطقس. تغطي قاعدة المعطيات التي تحتوي على عينات التدريب معلومات عن عوامل الطقس وقيمها المقابلة. ونظراً لأن عينات التدريب تكون غير كاملة ومن المحتمل غامضة نتيجة لمميزات التنبؤ بالطقس، لا يمكن أن تشتق قواعد القرار الدقيقة بواسطة الطرق القياسية. وهدفنا في هذا الفصل هو اقتراح طريقة لتوليد قواعد القرار (قواعد التصنيف) من المعلومات غير الكاملة. وتعتمد هذه الطريقة على الاستقراء الموجة بالصفة بعد الدراسة تأثير عناصر الطقس على بعضها البعض باستخدام نموذج تحليل الانحدار المشروح في الفصل الثالث وتمديد (توسيع) نموذج المجموعات التقريبية. استخدمت المعلومات الإحصائية لتعريف المناطق الموجبة والسالبة لمفهوم. توصف كل قاعدة تصنيف مولدة من قبل نظام التعلم بعامل تأكد، الذي يكون في الواقع احتمال أن الشيء المطابق للجزء الشرطي من قاعدة ينتمي إلى المفهوم.

في هذا الفصل، نقدم المفاهيم الأساسية لنظرية المجموعات التقريبية التي أصبحت نظرية رائجة في مجال التنقيب بالمعطيات. كما نقدم التنفيذ والتجارب لهذه الطريقة المقترحة.

ينظم هذا الفصل كما يلي: يقدم المقطع ٤-٢ نظرية المجموعات التقريبية ومفاهيمها الأساسية، ويقدم المقطع ٤-٣ التنفيذ والتجارب لنظام التعلم DBROUGH، و في المقطع ٤-٤ تم عرض النتائج التجريبية، ويختتم الفصل بمناقشه لمنهجنا في المقطع ٤-٥.

٤-٢ نظرية المجموعات التقريبية ومفاهيمها الأساسية

غالباً ما نجد أنفسنا في الكثير من المواقف، في حالة عدم تأكد. وهذا يمكن أن ينتج من الافتقار للمعرفة أو من عدم الكمال للمعطيات التي بين أيدينا. وكانت هذه الملاحظة الباعث خلف العديد من المحاولات من قبل الفلاسفة وعلماء المنطق لسنوات عديدة للتعامل مع مشاكل الغموض وعدم التأكد [٤٧].

لقد طبق علماء الحاسوب، على وجه الخصوص الباحثون المهتمون في الذكاء الصنعي، أفكار من حقول (مجالات) مختلفة على هذا المنطقة من البحث. وأهم هذه الحقول كان التحليل الإحصائي ونظرية المجموعات الضبابية ونظرية الدليل [٤٨]. وتتطلب كل هذه الطرق وسطاء من خارج الظاهرة المرصودة، أو تفرض مسبقاً أن الخواص تكون ذات وصف نوعي وتكون خاضعة لتأثير عشوائي لكي يكون بالإمكان تطبيق الطرق الإحصائية مثل تحليل التباين، أو الانحدار، أو الارتباط. ومؤخراً، لقد اكتسبت نظرية المجموعات التقريبية قبولاً. وتكون نظرية المجموعات التقريبية الخطوة الأولى وأحيانا الكافية في تحليل المعلومات غير كاملة أو غير المؤكدة. إنها تستخدم فقط المعلومات الداخلة ولا تعتمد على افتراض نموذج إضافي مثلما تفعل الطرق الأخرى. وهذا يعني بدلاً من استخدام الوسطاء الإضافية الأخرى، يستفيد تحليل المجموعة التقريبية بصورة فريدة من بنية المعطيات المعطاة [٤٦].

في المقاطع الفرعية التالية، نقدم المفاهيم الأساسية لنظرية المجموعات التقريبية.

٤-٢-١ نظم المعلومات

نفترض أن المجموعة المعطاة لعينات التدريب تمثل المعرفة حول النطاق. في منهجنا، توصف مجموعة التدريب بنظام المعلومات [٢٥]. تكون أنظمة المعلومات النماذج الرسمية لأنظمة الصفة– القيمة. لقد درست خواصها المنطقية بصورة موسعة من قبل العديد من الباحثين. وتكون المكونة الأساسية لنظام المعلومات مجموعة الأشياء. وتقليدياً، يمكن أن يمثل نظام المعلومات بجدول معطيات يدعى جدول المعلومات. وأعمدته تسمى بأسماء الصفات وسطوره تقابل أشياء المجموعة الشاملة. ويمثل كل سطر المعلومات عن شيء في هذه المجموعة الشاملة بلغة قيم الصفات. وتقليديا، تكون الصفات المستخدمة لوصف أشياء نظام المعلومات بعض السمات الابتدائية للأشياء[٤٩].

تعريف: رسمياً، يكون نظام المعلومات الرباعية $<U,A,V,f>$ حيث $V=\{x,x_2....x_n\}$ هـي مجموعـة منتهية غير خالية من الأشياء وA هي مجموعة منتهية غير خالية من الصفات. والصفات فـي A تصـنف أيضا إلى صفات شرطية C وصفة قرار D بحيـث أن $A=C\cup D$ و $C\cap D=\phi$ و V_α $V=\cup_{\alpha\in A}V_\alpha$ هـي مجموعة غير خالية من قيم الصفات و V_α ترمز إلى نطاق الصفة α و $f:UXA\to V$ هي دالة المعلومات التي تخصص قيمه للصفة $a\in A$ لكل شيء من U [٥٠].

يوضح مثال عن نظام المعلومات في الجدول ٤-١. تمثل السطور للجدول المسـماة x1,x2,….. xn، كما ذكرنا سابقا، الأمثلة (الأشياء). بينما تمثل الأعمدة قيم الصفات المنتمية إلى هذه الأشياء. في نظـام المعلومات، يميز نوعان من الصفات، الصفات الشرطية، وأيضا تدعى صفات التنبـؤ، وصـفات القـرار، وأيضا تدعى الصفات المتنبأ بها. وعادة، صفة قرار فردية يكون كل ما هو مطلوب.

Objects			C			D
	a0	a1	a2	a3	a4	a5
x1	Very cold	Clear	Normal	Calm	High	Moderate
x2	Cold	Partly	Moderate	Light air	Relatively high	Moderate
x3	Cold	Cloudy	Moderate	Light air	Relatively high	Moderate
x4	Cold	Fairly	Moderate	Light air	Relatively high	Moderate
x5	Very cold	Fairly	Normal	Calm	High	Moderate
x6	Very cold	Cloudy	Normal	Calm	Moderate	Moderate
x7	Cold	Clear	Normal	Calm	High	Moderate
x8	Warm	Partly	Relatively high	Light breeze	Relatively high	Moderate
x9	Warm	Partly	Relatively high	Light breeze	High	High
x10	Warm	Cloudy	Relatively high	Light breeze	High	High
x11	Warm	Partly	Moderate	Light breeze	High	High
x12	Warm	Cloudy	Moderate	Light breeze	High	High
x13	Very Cold	Partly	Normal	Calm	High	High
x14	Very cold	Cloudy	Normal	Calm	High	High
.						
.						

الجدول (٤-١) مجموعة اختبار (نظام معلومات) للتنبؤ بالطقس

في نظام المعلومات هذا، توجد عدة أشياء تمثل عينات التدريب (مجموعة الأمثلة)، التي تغطي معلومات عن عوامل الطقس وقيمها المقابلة. ويمكن أن تأتي مجموعة الأمثلة هذه من خبير نطاق أو من التاريخ الماضي لأنماط الطقس [١٢]. بفرض أننا نريد أن نكتشف القواعد التي تتنبأ بتوقع عنصر معين (على سبيل المثال، ضغط بخار الماء). وبناء على ذالك، يتم اختيار الصفة التي تمثل بخار الماء كصفة قرار، تدعى أيضا صفه تابعه، وبقية الصفات: درجة الحرارة،السماء، الرطوبة وسرعة الرياح والضغط الجوي تكون الصفات الشرطية، تدعى أيضا الصفات المستقلة. ويحتوي نظام المعلومات لمجموعة أمثلتنا للتنبؤ بالطقس على العديد من الأشياء، التي تشكل المعلومات عن الصفات الشرطية وصفة القرار. أربعة عشر شيئا من هذه الأشياء توضح في الجدول ٤-١.

في المقاطع ٤-٢-٢ و٤-٢-٣، سوف نستخدم الجدول ٤-١ كمثال لنظام المعلومات لتوضيح المصطلحات.

٤-٢-٢ تحليل المعطيات

٤-٢-٢-١ علاقة عدم إمكانية التمييز

يكون المفهوم الرئيسي لنظرية المجموعات التقريبية علاقة عدم إمكانية التمييز، المرتبطة عادة بمجموعة من الصفات. وإذا وجدت هذه العلاقة بين شيئين، هذا يعني أن كل قيم خواصها تكون متطابقة بالنسبة للصفات التي تحت الاعتبار ونتيجة لذلك هذين الشيئين لا يمكن أن يميزا بواسطة هذه الصفات.

تعريف: ليكن $S=<U,A,V,f>$ نظام معلومات ولتكن B مجموعة فرعية من A، وليكن xi وxj شيئان من U. تعرف العلاقة الثنائية، التي تدعى علاقة عدم إمكانية التمييز، بالصورة التالية [٥١. ٥٢]

$$IND(B)=\{(x_i,x_j)\in UXU/\forall\alpha\in B, f(x_i,\alpha)=f(x_j,\alpha)\}$$

نقول أن xi وxj غير قابلان للتمييز بواسطة مجموعة الصفات B في S= <U,A,V,f > إذا وفقط إذا f(xi, a)=f(xj, a) من أجل كل $\alpha\in B$. على سبيل المثال، في نظام المعلومات الموضح في الجدول ٤-١، تكون الأشياء x1, x2, x3, x4, x5, x6, x7, x8 غير قابلة للتمييز بواسطة مجموعة الصفات D.

تكون $IND(B)$ علاقة تكافؤ على U من أجل كل $B\subseteq A$. وهكذا، يمكن أن نعرف علاقتي تكافؤ طبيعيتين هي IND(C) وIND(D) على U من أجل نظام المعلومات S. ويكون المفهوم Y صف تكافؤ للعلاقة IND(D). وبدون فقدان للعمومية، يمكن أن نعتبر D كمجموعة مؤلفة من عنصر واحد. هدفنا أن

نبني قواعد القرار لكل مفهوم. وبمعرفة المفهوم Y، فإن التجزئة لـــ U بالنســـبة لهـــذا المفهـــوم، تعــرف بالصورة التالية:

$$\overset{*}{IND}(D) = \{Y, U - Y\} = \{Y, \neg Y\}$$

اعتمادا على مجموعة الصفات الشرطية C، يحدد الشيء xi صف تكافؤ $[x_i]_{IND}$ للعلاقة IND(C):

$$[x_i]_{IND} = \{x_j \in U / \forall \alpha \in C, f(x_j, \alpha) = f(x_i, \alpha)\}$$

تعريف: نقول أن $x_i \in U$ ينتمي بلا ريب إلى المفهوم y إذا $[x_i]_{IND} \subseteq Y$ وأن $x_i \in U$ من الممكن ينتمـــي إلى y إذا $[x_i]_{IND} \cap Y \neq \phi$ [٤٩].

تعريف: يعرف احتمال وقوع الحدث y مشروطا بوقوع الحدث $[x_i]_{IND}$ بالصورة التالية[٤٠]:

$$P(Y/[x_i]_{IND}) = \frac{p(Y \cap [x_i]_{IND})}{P([x_i]_{IND})} = \frac{card(Y \cap [x_i]_{IND})}{card([x_i]_{IND})}$$

هذا يعني، $P(Y/[x_i]_{IND}) = 1$ إذا وفقط إذا $[x_i]_{IND} \subseteq Y$ و $P(Y/[x_i]_{IND}) > 0$ إذا وفقط إذا $[x_i]_{IND} \cap Y \neq \phi$ و $P(Y/[x_i]_{IND}) = 0$ إذا وفقط إذا $[x_i]_{IND} \cap Y = \phi$

Algorithm (Concepts in Information System)

Input:
(1)The task relevant data relation
(2)The set D defining decision attributes, usually containing one attribute
Output:
Concepts in Information System
Method:
Apply an attribute-oriented regression analysis model induction algorithm.
Find the equivalence classes on the relation IND(D) according to

$Y_i = [x_i]_{IND} = \{x_j \in U | \forall a \in D, f(x_j, a) = f(x_i, a)\}$.

Algorithm (Conditional probability of the concept Y conditioned on event $[x_i]_{IND}$)

Input:
(1)The task relevant data relation.
(2)The set C defining condition attributes.

(3)The concept Y

Output:

Conditional probability of the concept Y conditioned on event $[x_i]_{IND}$

Method:

Apply an attribute-oriented regression analysis model induction algorithm.
Find the equivalence classes on the relation IND(C).
Apply the following law:

$$P(Y/[x_i]_{IND})=\frac{P(Y\cap[x_i]_{IND})}{P([x_i]_{IND})}=\frac{Card(Y\cap[x_i]_{IND})}{Card([x_i]_{IND})}$$

مثال ٤-٢-١

لنأخذ نظام المعلومات المعطى في الجدول ١. ٤. المفاهيم في نظام المعلومات هذا، أي صفوف التكافؤ على العلاقة IND(D) هي:

y0={x1,x2,x3,x4,x5,x6,x7,x8}
y1={x9,x10,x11,x12,x13,x14}

صفوف التكافؤ على العلاقة IND(C)هي:

$X1=[x1]_{IND}=\{x1\}$, $X2=[x2]_{IND}=\{x2\}$,
$X3=[x3]_{IND}=\{x3\}$, $X4=[x4]_{IND}=\{x4\}$,
$X5=[x5]_{IND}=\{x5\}$, $X6=[x6]_{IND}=\{x6\}$,
$X7=[x7]_{IND}=\{x7\}$, $X8=[x8]_{IND}=\{x8\}$,
$X9=[x9]_{IND}=\{x9\}$, $X10=[x10]_{IND}=\{x10\}$,
$X11=[x11]_{IND}=\{x11\}$, $X12=[x12]_{IND}=\{x12\}$,
$X13=[x13]_{IND}=\{x13\}$, $X14=[x14]_{IND}=\{x14\}$.

لاحظ أن x1,x2,x3,x4,x5,x6,x7. x8 تنتمي بلا ريب إلى المفهوم y0 لأن

$$x1, x2, x3, x4, x5, x6, x7, x8 \subseteq Y_0$$

الأشياء x9,x10,x11,x12,x13,x14 لا تنتمي إلى المفهوم y0لأن تقاطع صفوف تكافؤها مع y0 يساوي إلى الخالية. وأيضاً، الأشياء x9,x10,x11,x12,x13,x14 تنتمي إلى المفهوم y1 لأن

$$x9, x10, x11, x12, x13, x14 \subseteq Y_1$$

الأشياء x1,x2,x3,x4,x5,x6,x7,x8 من الممكن ألا تنتمي إلى المفهوم y1 لأن تقاطع صفوف تكافؤها مع y1 يساوي الخالية.

الاحتمالات الشرطية لـ y0 و y1 مشروطة بكل صف تكافؤ كما يلي:

P(y0/x1) = P(y0/x2) = P(y0/x3) = P(y0/x4) = P(y0/x5) = P(y0/x6) = P(y0/x7) = P(y0/x8) =1

P(y0/x9) = P(y0/x10) = P(y0/x11) = P(y0/x12) = P(y0/x13) = P(y0/x14) = 0
P(y1/x1) = P(y1/x2) = P(y1/x3)= P(y1/x4) = P(y1/x5)= P(y1/x6) = P(y1/x7) = P(y1/x8) = 0
P(y1/x9) = P(y1/x10)= P(y1/x11)= P(y1/x12) = P(y1/x13) = P(y1/x14) =1

٤-٢-٢-٢ تقريبات المجموعات

يكون العنصر الأولي(الأساسي) لنموذج المجموعات التقريبية فكرة فضاء التقريب [٤٩].

تعريف: من أجل نظام المعلومات $S =< U, A, V, F >$ وعلاقة التكافؤ IND(C) على U، يدعى الزوج المرتب $AS = (U, IND(C))$ فضاء التقريب بناءً على الصفات الشرطية C [٢٥].

تعريف: تدعى صفوف التكافؤ للعلاقة IND(C) المجموعات الابتدائية في AS لأنها تمثل أصغر مجموعات الأشياء التي تكون قابلة للتمييز بلغة الصفات وقيمها. ويدعى أي اجتماع منته من المجموعات الابتدائية في AS مجموعة محددة في AS [٤٩].

تعريف: لتكن $Y \subseteq U$ مجموعة فرعية من الأشياء التي تمثل مفهوم ولتكن $IND^*(C) = \{X_1, X_2, ...X_n\} = \{[x_1]_{IND}, [x_2]_{IND}, ..., [x_n]_{IND}\}$ مجموعة صفوف التكافؤ المحدثة بواسطة العلاقة IND(C). وفي نموذج المجموعات التقريبية القياسي، تعرف التقريبات السفلية والعلوية والمنطقة الحدية للمجموعة Y كما يلي [٤٧، ٥٠] :

(١) **التقريب السفلي لـِ** Y، يرمز له بالرمز $\underline{IND(C)}(Y)$ ، وهو اجتماع على المجموعات الابتدائية التي كل واحدة منها يمكن أن تصنف منتمية بلا ريب إلى المجموعة Y بناء على معلومات التصنيف الممثلة بعلاقة عدم إمكانية التمييز IND:

$$\underline{IND(C)}(Y) = \{x_i \in U / [x_i]_{ind} \subset y\} = U_{p(Y/X_i)=1}\{X_i \in IND^*(c)\}$$

(٢) **التقريب العلوي لـِ** Y، يرمز له بالرمز $\overline{IND(C)}(Y)$، هو اجتماع كل المجموعات الابتدائية التي كل واحدة منها يمكن أن تصنف منتمية من الممكن إلى المجموعة y اعتماداً على معلومات التصنيف الممثلة بواسطة علاقة عدم إمكانية التمييز IND:

$$\overline{IND(C)}(Y) = \{x_i \in U / [x_i]_{IND} \cap Y \neq \phi\} = U_{P(Y/X_i)>0}\{X_i \in IND^*(C)\}$$

(٣) **المنطقة الحدية**، يرمز لها بالرمز BND(Y)، وهي اجتماع كل المجموعات الابتدائية التي كل واحدة منها لا يمكن أن تحدد بالتأكيد ما إذا الأشياء من هذه المنطقة تنتمي إلى المجموعة Y أو متممة ¬Y:

$$BND(Y) = \overline{IND(C)}(Y) - \underline{IND(C)}(Y)$$

Algorithm (Lower and Upper Approximations)

Input:
(1)The task relevant data relation
(2)The set C defining condition attributes
(3)The set D defining decision attributes, usually containing one attribute
(4)The concept Y
Output:
The set of objects $\overline{IND(C)}(Y)$ or $\underline{IND(C)}(Y)$ defining the lower
Or upper approximation of y, respectively
Method:
Apply an attribute-oriented regression analysis model induction algorithm
Find IND*(C)={x1, x2, …, xn}={$[x1]_{IND}$, $[x2]_{IND}$, …, $[xn]_{IND}$}- The collection of equivalence classes induced by the relation IND(C)
Apply Algorithm (Conditional probability of the concept y conditioned on event $[xi]_{IND}$)
If P(y/xi) = 1 then $\underline{IND(C)}(Y) = \cup$ {xi ∈ IND*(C)}
If P(y/xi) > 0 then $\overline{IND(C)}(Y) = \cup$ {xi ∈ IND*(C)} [49]

Algorithm (Boundary Region)

Input:
(1)The task relevant data relation
(2)The set C defining condition attributes
(3)The concept Y
Output:
The set of objects BND(Y)- Boundary region of Y.
Method:
Apply an attribute-oriented regression analysis model induction algorithm
Find $\overline{IND(C)}(Y)$ and $\underline{IND(C)}(Y)$ based on Algorithm (Lower and Upper Approximations)

$$BND(Y) = \overline{IND(C)}(Y) - \underline{IND(C)}(Y) \quad [49]$$

٤-٢-٢-٣ التصنيف التقريبي الاحتمالي – β

لا تستفيد التعاريف أعلاه من المعلومات الإحصائية في المنطقة الحدية BND(y). لهذا السبب، تم اقتراح عدد من التوسعات (التمديدات)على نموذج المجموعات التقريبية الأصلي [٥٣، ٥٦]. في منهجنا، نحاول أن نصحح القيد بإدخال فضاء التقريب – β.

تعريف: يكون فضاء التقريب الاحتمالي – β، الذي يرمز له بالرمز AS_P، الثلاثية $< U, IND(c), p >$، حيث P القياس الاحتمالي الموصوف في المقطع الفرعي (٤-٢-٢-١) و β هو عدد حقيقي في المجال [0. 25,1] [٤٩]. ويمكن أن يقسم فضاء التقريب الاحتمالي – β إلى منطقتين: المنطقة الموجبة – β والمنطقة السالبة – β.

تعريف: تكون المنطقة الموجبة – β للمجموعة Y، الذي يرمز لها بالرمز POSc(Y)، كل المجموعات الابتدائية لـ U التي يمكن أن تصنف في المفهوم Y باحتمال شرطي $P(Y/x_i)$ أكبر أو يساوي إلى الوسيط β [٤٩]:

$$\mathrm{POS_C}(Y) = \bigcup P(Y / X_i) \geq \beta, \{X_i \in IND^*(C)\}$$

تعريف: تكون المنطقة السالبة – β للمجموعة Y، الذي يرمز لها بالرمز NEGc(Y)، كل المجموعات الابتدائية لـ U التي يمكن أن تصنف في المفهوم $\neg Y$ باحتمال شرطي $P(y/x_i)$ أصغر من الوسيط β [٤٩]:

$$\mathrm{N}EG_{\mathrm{C}}(Y) = \bigcup P(Y / X_i) < \beta, \{X_i \in IND^*(C)\}$$

تعريف: ليكن $X_i \in Y$ شيئاً ولتكن POSc(Y) وNEGc(Y) المناطق الموجبة والسالبة للمفهوم Y على التوالي. يصنف الشيء xi منتميا إلى المفهوم Y إذا وفقط إذا $xi \in POSc(Y)$ أو منتميا إلى المتممة $\neg Y$ للمفهوم y إذا وفقط إذا $xi \in NEGc(Y)$ [٤٩].

في الواقع نريد أن نقرر ما إذا xi يكون في المفهوم Y. على أساس مجموعة صفوف التكافؤ في AS_P وليس على أساس المجموعة Y. هذا يعني، أننا نتعامل مع $POSc(Y)$ و $NEGc(Y)$ بدلاً من المجموعة Y.

مثال: ٤-٢-٢

من المثال **٤-٢-١**،تذكر $Y0=\{x1,x2,x3,x4,x5,x6,x7,x8\}$ و $Y1=\{x9,x10,x11,x12,x13,x14\}$

لإيجاد المناطق الموجبة والسالبة y0، نتبع الأتي:

$IND^*(C) =\{\{x1\}, \{x2\}, \{x3\}, \{x4\}, \{x5\}, \{x6\}, \{x7\}, \{x8\}, \{x9\}, \{x10\}, \{x11\}, \{x12\}, \{x13\}, \{x14\}\}$

$P(y0/x1) = P(y0/x2) = P(y0/x3) = P(y0/x4) = P(y0/x5) = P(y0/x6) = P(y0/x7) = P(y0/x8) =1$

$P(y0/x9) = P(y0/x10) = P(y0/x11) = P(y0/x12) = P(y0/x13) = P(y0/x14) = 0$

Let β=0. 4. Therefore,

$POSC(y0) = \cup P(y1/xi) \geq \beta, \{ xi \in IND^*(C)\}=\{x1, x2, x3, x4, x5, x6, x7, x8\}$

$NEGC(y0)= \cup P(y1/xi)<\beta, \{ xi \in IND^*(C)\}=\{x9, x10, x11, x12, x13, x14\}$

وبالمثل، يمكن أن نوجد المناطق الموجبة والسالبة للمفهوم Y1.

٤-٢-٤ اختزال الصفات الشرطية

إن إحدى المشاكل الأساسية، التي يمكن أن تحل باستخدام منهج المجموعات التقريبية، هي اختزال الصفات من قاعدة المعطيات، أو كما يسمى في سياق المجموعة التقريبية، نظام المعلومات. في نظام المعلومات، توجد غالباً بعض الصفات الشرطية التي لا تقوم بإسهام معنوي في تعريف الأشياء في U. هذا يعني، إنها لا تقدم أي معلومات إضافية حول الأشياء في U. إنه من المرغوب فيه أن نزيل تلك الصفات لأن درجة تعقيد وتكلفة عملية القرار يمكن أن تقلل (تخفض) إذا أزيلت هذه الصفات الشرطية [٤٩، ٥٥]. في هذا المقطع الفرعي، سنستخدم مفهوم المختزل في المجموعات التقريبية لوصف طريقة لاختزال الصفات الشرطية.

تعريف: ليكن $S =< U, C \cup \{d\}, V, f >$ نظام معلومات، يقال عن الصفة α أنها غير ضرورية في C بالنسبة إلى{d} إذا كان $POS_C(Y)=POS_{C-\{\alpha\}}(Y)$ وإلا تكون الصفة α ضرورية في C بالنسبة إلى {d} [٢٥].

تعريف: يقال عن المجموعة الفرعية من الصفات الشرطية $B \subseteq A$ أنها مجموعة تابعة في S بالنسبة إلى {d} إذا وجدت مجموعة فرعية $K \subset B$ بحيث أن $POS_B(y)=POS_K(y)$ وإلا تكون B مجموعة مستقلة بالنسبة إلى{d}.

تعريف: يكون المختزل $C^{'}$ للصفات الشرطية C هو أكبر مجموعة فرعية مستقلة من الصفات الشرطية بالنسبة إلى {d} [٢٥].

يكون الإجراء لإيجاد مختزل فردي صريح جداً. نأخذ الصفة الشرطية $a \in C$. إذا POSC-{α}(y)=POSC(y)، عندئذ الصفة a تعلم بأنها مكررة **(**زائده**)** وتزال من مجموعة الصفات الشرطية C. ويمكن أن تزال الصفات الشرطية الزائدة الأخرى بنفس الطريقة. وتكون مجموعة الصفات الشرطية المتبقية هي المختزل.

Algorithm (Reduct $C^{'}$ of condition attributes C)

Input:

(1)The task relevant data relation.

(2)The set C defining condition attributes.

(3)The set D defining decision attributes, usually containing one attribute.

(4)The attribute a.

Output:

The set of attributes representing reduction of condition attributes.

Method:

Apply an attribute-oriented regression analysis model induction algorithm.

Find $POS_C(y)$.

Find $POS_{C-\{\alpha\}}$ (y), $\forall$ a$\in$C.

If $POS_C(y)$= $POS_{C-\{\alpha\}}$(y), for some a$\in$C, then RED = C-{a} [1].

مثال ٤-٢-٣

في هذا المثال سنوجد المختزل للصفات الشرطية. يكون الإجراء لإيجاد مختزل فردي كما يلي:

المنطقة الموجبة لـ y1={x9,x10,x11,x12,x13,x14} بالنسبة للصفات C هي:

POS_C (y1)={x9,x10,x11,x12,x13,x14}

المنطقة الموجبة لـ y1 بالنسبة إلى $C-\{\alpha_0\}$ هي:

$POS_{C-\{\alpha 0\}}$ (y1)={x9,x10,x11,x12,x13,x14}

بما أن $POS_{C-\{\alpha 0\}}(Y1) \neq POS_C(Y_1)$ ، لذلك يمكن إزالة α_0 .

المنطقة الموجبة لـ y1 بالنسبة إلى $C-\{\alpha_1\}$ هي:

$POS_{C-\{\alpha 1\}}$(y1) ={x9,x10,x11,x12,x13,x14}

بما أن $POS_{C-\{\alpha1\}}(Y1) \neq POS_C(Y_1)$، لذلك لا يمكن إزالة α_1.

المنطقة الموجبة إلى $C-\{\alpha_2\}$ هي:

$POS_{C-\{\alpha2\}}$ (y1) ={x9,x10,x11,x12,x13,x14}

بما أن $POS_C(Y_1)$= $POS_{C-\{\alpha2\}}$ (y1)، لذلك يمكن إزالة α_2 .

المنطقة الموجبة لـِ y1بالنسبة إلى $C-\{\alpha_3\}$ هي:

$POS_{C-\{\alpha3\}}$ (y1) ={x9,x10,x11,x12,x13,x14}

بما أن $POS_{c-\{\alpha3\}}(y_1) = POS_c(Y)$ لذلك يمكن إزالة $\alpha3$.

المنطقة الموجبة لـِy1 بالنسبة إلى $C-\{\alpha4\}$ هي:

$POS_{C-\{\alpha4\}}(Y_1)$={x6,x8,x9,x10,x11,x12,x13,x14}

بما أن $POS_{C-\{\alpha4\}}(Y_1) \neq POS_C(Y_1)$، لذلك لا يمكن إزالة $\alpha4$.

ونتيجة لذلك، نستنتج أن المختزل للصفات الشرطية هو: RED ={ α_1, α_4}

٤-٢-٣ توليد قواعد القرار

يعتبر توليد قواعد القرار أحد أهم المشاكل التي يمكن أن تقرب باستخدام نظرية المجموعات التقريبية. في سياق اكتشاف المعرفة أو التنقيب بالمعطيات، عادة تكون مجموعة معطياتنا قاعدة معطيات علائقية [٥٠].

يمكن أن تعتبر قاعدة المعطيات العلائقية كنظام معلومات الذي فيه الأعمدة تسمى بالصفات والأسطر تسمى بالأشياء، والمدخل في العمود α والسطر xi يملك القيمة f(xi, a). تشكل مجموعة كل الحدوديات مجموعة عينة التدريب.

في عملية توليد القاعدة، تصنف الصفات المعتبرة في مجموعتين منفصلتين: الصفات الشرطية وصفات القرار. عادة تكون صفة القرار الفردية كل ما هو مطلوب وتكون قاعدة القرار في نظام المعلومات أو في قاعدة المعطيات العلائقية خليط من قيم بعض الصفات الشرطية مطبقة إلى قيمة خاصة لصفة القرار [٢٥]:

$$r: \text{Condition}(r) \rightarrow \text{Decision}(r)$$

وعادة تكون المدخلات لعملية توليد القاعدة مجموعة من الصفات الشرطية C وصفة القرار {d}.

في هذا المقطع، سنصف كيف تولد قواعد القرار بناء" على المختزل.

لتكن $IND^*(RED)=\{X_1,X_2......X_n\}$ مجموعة صفوف التكافؤ للعلاقة $IND(RED)$ حيث RED هو المختزل الذي هو مجموعة مختزلة من الصفات الشرطية C في S. ولتكن $IND^*(D)=\{Y,\neg Y\}$ التجزئة المحدثة بواسطة صفة القرار. يرتبط كل صف تكافؤ xi لعلاقة التكافؤ $IND(RED)$ مع خليط فريد من قيم الصفات المنتمية إلى RED. ويشار إلى هذا الخليط من القيم بتوصيف صف التكافؤ $X_i \in IND^*(RED)$. ويمكن أن يعبر عن توصيف صف التكافؤ xi بالصورة التالية:

$$Des(X_i)=\wedge_{\alpha\in RED}(\alpha=f(x_i,\alpha))$$

حيث يرمز $\wedge$ إلى مؤثر الوصل وxi هو شيء في صف التكافؤ xi. بالمثل، تكون التوصيفات لـ y و $\neg Y$ بالشكل:

$$Des(Y) = (D=f(xi,D))$$

$$Des(\neg Y)=(D\neq f(x_i,D))$$

حيث D هي صفة القرار و $x_i \in Y$.

يمكن أن توصف العلاقة بين التجزئة IND*(RED) والتجزئة $IND^*(D)$ بقواعد القرار الآتية:

من أجل $X_i \in IND^*(RED)$

$$Des(X_i)\xrightarrow{ci} Des(Y),\ if\ P(Y/X_i)\geq\beta$$

$$Des(X_i)\xrightarrow{ci} Des(\neg Y),\ if\ P(Y/X_i)<\beta,$$

حيث ci عامل التأكد [٤٢]: Ci=max(P(Y/Xi),1-P(Y/Xi)) هذا يعني أنه إذا شيء xi حقق التوصيف Des(xi) وإذا $P(Y/Xi)\geq\beta$، عندئذ الشيء xi ينتمي إلى y بعامل التأكد Ci. بصورة مشابه، إذا P (y/xi)<β، عندئذ الشيء xi من الممكن ينتمي إلى المتممة $\neg Y$ بعامل تأكد Ci.

Algorithm (Decision Rule Generation)

Input:

(1)The task relevant data relation .

(2)The set C defining condition attributes .

(3)The set D defining decision attributes, usually containing one attribute

(4)The concept y

Output:

The set of generated decision rules for the information system .

Method:

Apply an attribute-oriented regression analysis model induction algorithm .

Find RED according to algorithm (Reduct C' of condition attributes C) .

Find IND*(RED)- the collection of equivalence classes of the relation IND(RED) .

Find the descriptions of equivalence classes according to the following:

$Des(xi)=\wedge a \in RED\ (a=f(xi, a)),\ xi \in xi$.

Find the descriptions of y and $\neg y$ according to the following:

$Des(y)=(D=f(xi, D))$ and $Des(\neg y)=(D \neq f(xi, D)),\ xi \in y$.

Find the conditional probabilities P(y/xi) according to algorithm (Conditional probability of the concept y conditioned on event [xi]IND) .

If $P(y/xi) \geq \beta$, then $Des(xi) \xrightarrow{ci} Des(y)$.

If $P(y/xi) < \beta$, then $Des(xi) \xrightarrow{ci} Des(\neg y)$.

Where ci is the certainty factor: ci =max(P(y/xi),1- P(y/xi)) [49]

مثال ٤-٢-٤

في هذا المثال، نوجد قواعد القرار بالنسبة للمفهوم y0 والمفهوم y1. سيكون الإجراء لإيجاد قواعد القـــرار كما يلي:

من المثال٤-٢-٣، وجدنا RED={a1,a4} هو المختزل للصفات الشـــرطية. المجموعـــة IND*(RED) لصفوف التكافؤ للعلاقة IND (RED)هي:

IND*(RED)={{x1, x7}, {x2, x8}, {x3}, {x4}, {x5}, {x6}, {x9, x11, x13}, {x10, x12, x14}}.

التوصيفات لصفوف التكافؤ أعلاه هي:

Des(x1)= (a1=Clear)$\wedge$(a4=High)
Des(x2)= (a1=Partly)$\wedge$(a4=Relatively high)
Des(x3)= (a1=Cloudy)$\wedge$(a4=Relatively high)
Des(x4)= (a1=Fairly)$\wedge$(a4=Relatively high)
Des(x5)= (a1=Fairly)$\wedge$(a4=High)
Des(x6)= (a1=Clear)$\wedge$(a4=High)
Des(x7)= (a1=Partly)$\wedge$(a4=High)
Des(x8)= (a1=Cloudy)$\wedge$(a4=High)

التوصيفات لــ y0 و¬y0 و y1 و ¬y1 هي:

Des(y0)=(D=Moderate)
Des($\neg$y0)=(D$\neq$Moderate)
Des(y1)=(D=High)
Des($\neg$y1)=(D$\neq$ High)

الاحتمالات الشرطية لـِ y0 و y1 مشروطة بالحدث xi هي:

$P(y0/x1)= P(y0/x2)= P(y0/x3)= P(y0/x4)= P(y0/x5)= P(y0/x6)=1$.

$P(y0/x7)= P(y0/x8)=0$.

$P(y1/x1)= P(y1/x2)= P(y1/x3)= P(y1/x4)= P(y1/x5)= P(y1/x6)=0$.

$P(y1/x7)= P(y1/x8)=1$.

قواعد القرار بالنسبة للمفهوم y0 و y1 هي:

$(a1=Clear)\wedge(a4=High)\xrightarrow{1}(D=Moderate)$

$(a1=Partly)\wedge(a4=Relatively\ high)\xrightarrow{1}(D=Moderate)$

$(a1=Cloudy)\wedge(a4=Relatively\ high)\xrightarrow{1}(D=Moderate)$

$(a1=Fairly)\wedge(a4=Relatively\ high)\xrightarrow{1}(D=Moderate)$

$(a1=Fairly)\wedge(a4=High)\xrightarrow{1}(D=Moderate)$

$(a1=Cloudy)\wedge(a4=Moderate)\xrightarrow{1}(D=Moderate)$

$(a1=Partly)\wedge(a4=High)\xrightarrow{1}(D\neq Moderate)$

$(a1=Cloudy)\wedge(a4=High)\xrightarrow{1}(D\neq Moderate)$

$(a1=Clear)\wedge(a4=High)\xrightarrow{1}(D\neq High)$

$(a1=Partly)\wedge(a4=Relatively\ high)\xrightarrow{1}(D\neq High)$

$(a1=Cloudy)\wedge(a4=Relatively\ high)\xrightarrow{1}(D\neq High)$

$(a1=Fairly)\wedge(a4=Relatively\ high)\xrightarrow{1}(D\neq High)$

$(a1=Fairly)\wedge(a4=High)\xrightarrow{1}(D\neq High)$

$(a1=Cloudy)\wedge(a4=Moderate)\xrightarrow{1}(D\neq High)$

$(a1=Partly)\wedge(a4=High)\xrightarrow{1}(D=High)$

$(a1=Cloudy)\wedge(a4=High)\xrightarrow{1}(D=High)$

٤-٣ التنفيذ والتجارب لنظام التعلم DBROUGH

لاختبار وتجريب خوارزميات تعلم قاعدة المعطيات العلائقية المطورة السابقة، بني نظام تعلم قاعدة معطيات تجريبي، DBROUGH، لاكتشاف القواعد من المعطيات للتنبؤ بالطقس ضمن إطار عمل نظرية المجموعات التقريبية الموجه بالصفة ونموذج تحليل الانحدار.

٤-٣-١ معمارية نظام DBROUGH

يكون نظام DBROUGH نظام متكامل لتعميم المعطيات وتحليلها وتوليد القواعد ويستغل المفاهيم المختلفة في نظرية المجموعات التقريبية الموجه بالصفة ونموذج تحليل الانحدار لاكتشاف المعرفة من قواعد المعطيات العلائقية. نفذ نظام DBROUGH باستخدام لغة البرمجة Microsoft Visual C# under windows. وتوضح المعمارية لنظام DBROUGH في الشكل ٤-١. ويشتمل النظام على ثلاثة مكونات: قاعدة المعطيات العلائقية والنواة الحسابية، والسطح البيني لنظام DBROUGH. وتحتوي قاعدة المعطيات العلائقية على معطيات الطقس المرصودة مسبقاً عن طقس المدن السورية. وتعتبر النواة الحسابية قلب النظام والتي تكون مسؤولة عن وظائفه. وإنها تشتمل على البرامج الرئيسية لنظام DBROUGH المطورة لتعميم المعطيات وتحليلها وتوليد القواعد ضمن إطار عمل نظرية المجموعات التقريبية الموجه بالصفة. ويكون السطح البيني للنظام مكونة رئيسية أخرى في النظام المنفذ والتي تكون مسؤولة عن استخدامات النظام. ويكون السطح البيني لنظام DBROUGH سطح بيني رسومي للمستخدم (GUI) يتم تنفيذه في بيئة نظام التشغيل Windows. ويعكس تصميم السطح البيني لنظام DBROUGH محتويات نواة نظام DBROUGH ويسمح لوظائفه لأن توسع بطريقة بسيطة.

٤-٣-٢ وظيفة النظام

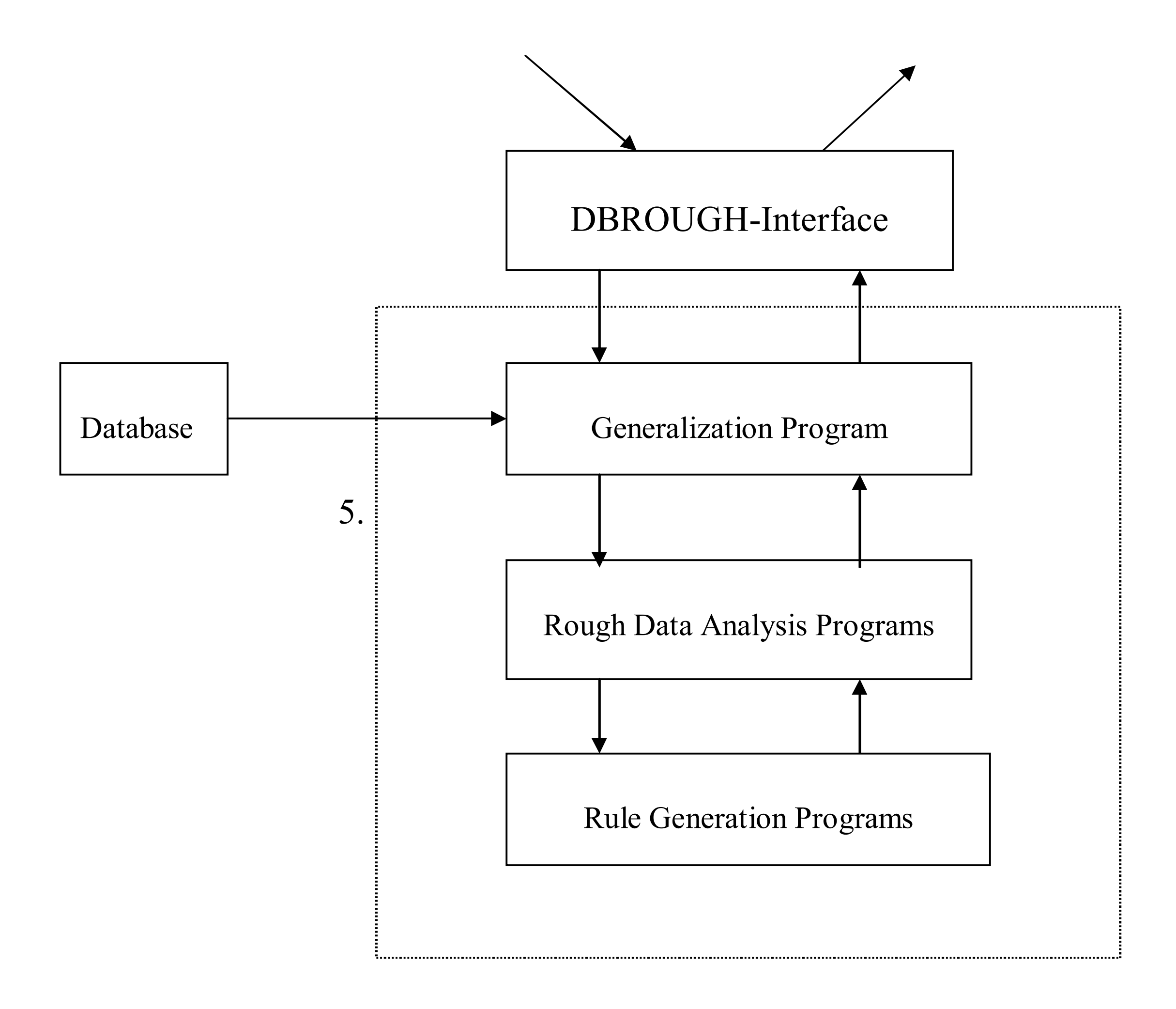
كما وصف سابقاً، تكون النواة الحسابية مسؤولة عن كل وظائف النظام. وتشمل نواة النظام على ثلاثة مجموعات رئيسية من البرامج والتي ترى كنظم فرعية لنظام DBROUGH. في هذا المقطع الفرعي، نلخص الوظائف الرئيسية للنظم الفرعية الثلاثة.

٤-٣-٢-١ نظام التعميم الفرعي

يكون نظام التعميم الفرعي مسؤولاً عن تنفيذ خوارزمية استقرار نموذج تحليل الانحدار التي خوارزمياتنا الرئيسية تعتمد عليها لتحليل المعطيات في سياق نظرية المجموعات التقريبية. وتعرض الوظائف الرئيسية لهذا النظام الفرعي أدناه:

- اختيار سجلات الطقس ذات الصلة بالمدينة والفصل.
- تنفيذ نموذج تحليل الانحدار لعنصر طقس معين بدلالة العناصر الأخرى.
- إزالة الصفة التي لا تملك مفهوم عالي المستوى في هرمية المفهوم.
- دمج الحدوديات المتطابقة في التعميم.
- تقديم تقديرات للاعتقاد في التعميم باستخدام المنطق الضبابي.

User Request   Learning Results



الشكل ٤-١ معمارية النظام DBROUGH

٤-٣-٢-٢ نظام تحليل المعطيات التقريبية الفرعي

يكون نظام تحليل المعطيات التقريبية الفرعي مسؤول عن المهام الرئيسية للنظام التي تعتمد عليها الخوارزميات الرئيسية لتوليد القواعد في سياق نظرية المجموعات التقريبية. وتعرض المهان الرئيسية لهذا النظام الفرعي أدناه:

* تحديد المفاهيم في نظام المعلومات.
* حساب الاحتمال الشرطي للمفهوم y مشروطاً بالحدث x.
* تحديد الصفات الشرطية وصفات القرار.
* حساب المناطق الموجبة والسالب لمفهوم.
* حساب المختزل للصفات الشرطية، أي، استخراج السمات الملائمة من المجموعة الكاملة للسمات المتاحة.

٤-٣-٢-٣ نظام توليد القواعد الفرعي

يكون نظام توليد القواعد الفرعي مسؤولاً عن توليد قواعد المعرفة من مجموعة المعطيات التي في المتناول. وتعرض المهام الرئيسية لهذا النظام الفرعي أدناه:

*- حساب توصيف صف التكافؤ x_i المنتمي إلى $IND^*(RED)$.

*- حساب توصيف المفهوم y والمفهوم ¬y المنتمي إلى التجزئة $IND^*(D)$.

*- إيجاد قواعد القرار الاحتمالية.

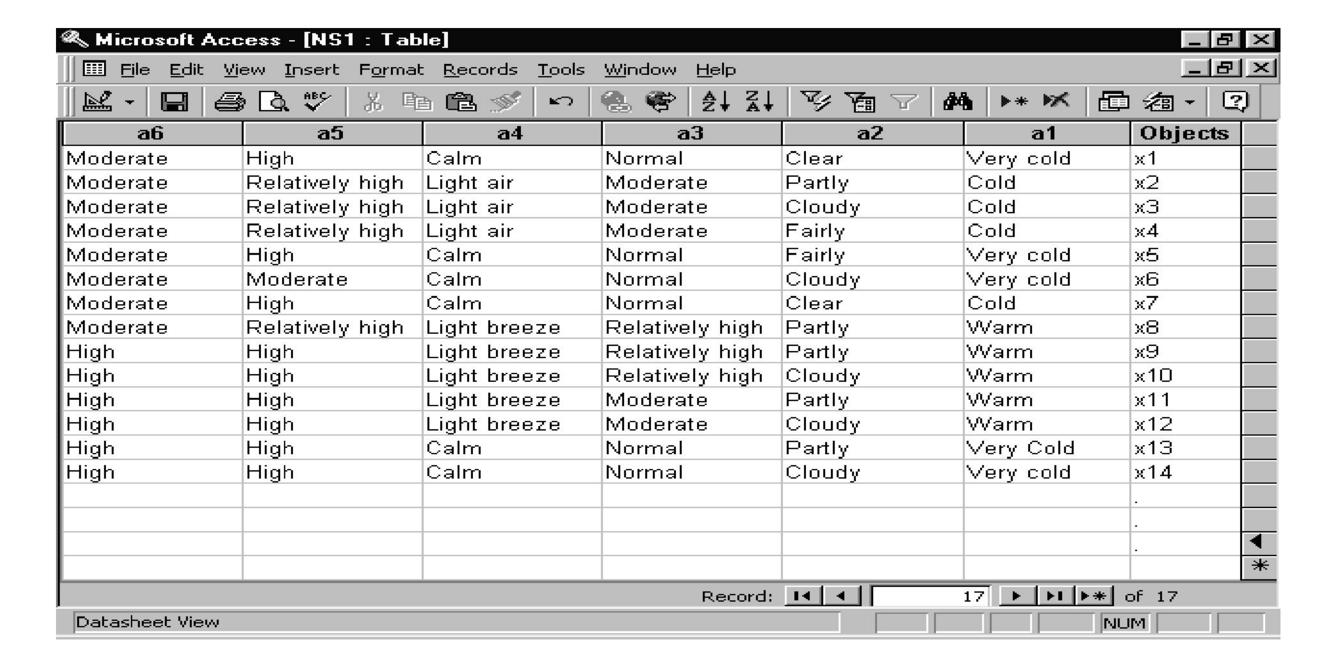
*- حساب عامل التأكد للقواعد.

٤-٣-٣ شكل الملفات

في هذا المقطع الفرعي، نصف أشكال الملفات المختلفة المستخدمة في نظام DBROUGH. يوجد ملفان الذي معهما يتصل النظام: ملف المعطيات الذي يحتوي على مجموعة المعطيات المختارة والتي عليها ستنفذ برامج التعميم وتحليل المعطيات وتوليد القواعد، وملف الخرج الذي إليه يوجه النظام نتائجه.

٤-٣-٣-١ ملف المعطيات

في شكله الحالي، يعمل النظام على المعطيات المزودة في شكل ملفات ODBC. ويكون شكل ملف المعطيات مشابهاً للشكل المستخدم من قبل ملفات ACCESS. ويوضح مثال لهذا الملف، من أجل نظام المعلومات في الجدول ٤-١، في الشكل ٤-٢. وتكون المعطيات محتواة في حدوديات. ويمثل العمود الأول أشياء نظام المعلومات. ويمثل العمود الأخير صفة القرار لنظام المعلومات. وتمثل الأعمدة المتبقية الصفات الشرطية.



Microsoft Access - [NS1 : Table]

File Edit View Insert Format Records Tools Window Help

a6	a5	a4	a3	a2	a1	Objects
Moderate	High	Calm	Normal	Clear	Very cold	x1
Moderate	Relatively high	Light air	Moderate	Partly	Cold	x2
Moderate	Relatively high	Light air	Moderate	Cloudy	Cold	x3
Moderate	Relatively high	Light air	Moderate	Fairly	Cold	x4
Moderate	High	Calm	Normal	Fairly	Very cold	x5
Moderate	Moderate	Calm	Normal	Cloudy	Very cold	x6
Moderate	High	Calm	Normal	Clear	Cold	x7
Moderate	Relatively high	Light breeze	Relatively high	Partly	Warm	x8
High	High	Light breeze	Relatively high	Partly	Warm	x9
High	High	Light breeze	Relatively high	Cloudy	Warm	x10
High	High	Light breeze	Moderate	Partly	Warm	x11
High	High	Light breeze	Moderate	Cloudy	Warm	x12
High	High	Calm	Normal	Partly	Very Cold	x13
High	High	Calm	Normal	Cloudy	Very cold	x14

Record: 17 of 17

Datasheet View NUM

الشكل ٤ـ٢ ملف معطيات نظام المعلومات للطقس

٤-٣-٣-٢ ملف الخرج

يكون ملف الخرج لنظام DBROUGH ملف ويندوز بسيط الذي إليه يوجه النظام القواعد المستقرأة. ويوضح الشكل لهذا الملف في الشكل ٤-٣. وتكون القاعدة محتواة في شكل مفصول فراغياً. وتفصل كل قاعدة عن القواعد الأخرى بسطر جديد ويعطى عدد القواعد في نهاية الملف.

[Rule1]
$(a_1=\text{Clear})\wedge(a_4=\text{High}) \xrightarrow{1} (D=\text{Moderate})$
[Rule2]
$(a_1=\text{Partly})\wedge(a_4=\text{High}) \xrightarrow{1} (D=\text{High})$
[TOTALRULES]

Number Of Rules=2

الشكل ٤ـ ٣ ملف الخرج

٤-٤ النتائج التجريبية

عندما نأتي إلى اختبار فعالية خوارزميات التعلم المشروحة في الفصول السابقة، تختبر عادة مقابل مجموعات معطيات العالم الحقيقي. ونظراً لأن المجال الذي نتعامل معه يكون مجال (نطاق) مناخي، لذلك نختبر طريقتنا المقترحة في ذلك النطاق. كما ذكر سابقاً، نفذت الطريقة المعروضة في الفيجوال #C. ينجز

نظام DBROUGH المهام المتعددة لتعميم المعطيات وتحليلها وتوليد القواعد. في هذا المقطـــع، عرضـــنا بعض التجارب لاختبار وظائف نظامنا لاكتشاف القواعد من المعطيات للتنبؤ الشهري بـــدرجات الحـــرارة، والسحب، والرطوبة النسبية، وسرع الرياح، والضغوط الجوية، وضغوط بخار الماء اعتماداً علـــى محطـــة طقسية واحدة.

التجربة على تطبيق مناخي:

في هذا التطبيق، عرضنا بعض التجارب لنظامنا على قاعدة معطيات طقسية حقيقية في مجالنـــا المنـــاخي، حيث يتعامل هذا المجال مع درجات الحرارة وكميات السحب، والرطوبة النسبية، وسرع الرياح، والضغوط الجوية، وضغوط بخار الماء. ويكون الهدف الرئيسي من التجارب المنفذة إثبات أن إطار العمـــل المقتـــرح يكون قابل للتطبيق لاكتشاف القواعد من المعطيات للتنبؤ بالطقس. أولاً، عرضنا وصف لقاعدة المعطيـــات الطقسية المستخدمة في هذا التطبيق. ثانياً، أجريت بعض التجارب على الوظائف المختلفة للنظام ومن ثم تم سرد النتائج.

٤-٤-١ وصف قاعدة المعطيات الطقسية

تكون قاعدة المعطيات الطقسية المستخدمة في هذا التطبيق قاعـــدة معطيـــات المعـــدل المنـــاخي الشـــهري المحصول عليها من قبل هيئة الأرصاد الجوية السورية. وتخزن قاعـــدة المعطيـــات الطقســـية المعطيـــات المناخية لثلاثة وعشرون سنة مرصودة مسبقاً من عام ١٩٨٥ إلى عام ٢٠٠٧ للمحطات الطقسية المبعثـــرة في كل أنحاء الجمهورية العربية السورية.

المعطيات المناخية :

تقسم المعطيات في قاعدة المعطيات الطقسية إلى ١٤ مجموعة معطيات. تصف كل مجموعـــة المعطيـــات المناخية في واحدة من المحافظات ١٤ الموجودة في الجمهورية العربية السورية. وتحتوي المعطيات على المعدل الشهري لدرجات الحرارة، وكميات السحب، والرطوبة النسبية. وسرع الرياح، والضغوط الجويـــة، وضغوط بحار الماء لهذه المحافظات في كل الفصول. وبغرض البساطة، ســـنفرض أن قاعـــدة المعطيـــات الطقسية تحتوي فقط على معطيات حمص للشهر الأول.

طبيعة المعطيات :

بكلام عام، تقسم الصفات التي تنتج نظام المعلومات إلى صفات وصفية وأخرى كمية. يكون المجال الأصلي للصفة الكمية عادةً مجموعة فرعية (أي فترة) من المحور الحقيقي. بينما يكون المجال للصفة الوصفية مجموعة منتهية من المصطلحات الوصفية، وعادة بقدرة منخفضة.

في قاعدة المعطيات الطقسية المختارة، تمثل الصفات القيم الحقيقية لدرجة الحرارة مقاسة في الدرجات المئوية، وكميات السحب مقاسة بالثمن، والرطوبة النسبية مقاسة بالمئة. وسرع الرياح مقاسة بالعقدة، والضغط الجوي مقاساً بالميلي بار، وضغط بخار الماء مقاساً بالميلي بار. وبناءً على ذلك، تمثل الصفات لمعطياتنا في شكل كمي.

فصل المعطيات :

عادةً، عندما تحتوي مجموعات المعطيات المعالجة على معطيات حقيقية، وعلى وجه الخصوص مع المعطيات الكمية، يكون من المحتمل عدد القيم لكل صفة كبير جداً. وهكذا، من المحتمل جدا، سيقود توليد القواعد لهذه المجموعاتً إلى قواعد تكون خاصة جداً بمجموعات المعطيات هذه. وعندما تصنف مجموعات معطيات مشابهة، ستكون درجة القيم للصفات غير ممثلة في القواعد عالية، وهذا يقود إلى تصنيف ضعيف. وهكذا في التطبيقات العملية، نادراً ما تستخدم مباشرةً قيم الصفات الكمية في تحليل المجموعات التقريبية. وبدلاً من ذلك، يقسم المجال إلى بضع فترات فرعية تمثل المصطلحات الوصفية.

من أجل قاعدة معطياتنا الطقسية، ومن خلال استخدام خوارزمية الاستقراء الموجه بالصفة المقدمة في الفصل الثالث، تقطع (تفصل) المعطيات الأصلية باستبدال القيم للصفات المختلفة بنطاق من الرموز (أي، مفاهيم عالية المستوى) تمثل مميزات المجال الذي تحت الدراسة. وهكذا نحصل على مجموعة المعطيات (أي نظام المعلومات) المستخدمة في هذا التطبيق والتي ستشرح في الفقرة القادمة.

حددت المفاهيم العالية المستوى بواسطة خبير طقس. ووفقاً لرؤية الخبير، الطريقة الملائمة أكثر لتخصيص رموز المجال (أي المفاهيم العالية المستوى) لكل صفة وضحت في الفصل الثالث.

نظام المعلومات المستخدم :

يتألف نظام المعلومات الذي عليه التجارب ستنجز من ٢٠ شيئاً بصفات تمثل حالات الطقس (أي أمثلة التدريب) لنظام معلوماتنا. وتوضح عشرة من هذه الأشياء في الجدول ٤-٤.

a6	a5	a4	a3	a2	a1	Object
high	relatively	moderate	moderate	Partly	slightly cold	x1
high	relatively	moderate	moderate	Fairly	slightly cold	x2
high	relatively	moderate	normal	Partly	warm	x3
moderate	high	moderate	normal	Fairly	slightly cold	x4
high	relatively	moderate	relatively	Fairly	slightly cold	x5
moderate	relatively	moderate	normal	Fairly	slightly cold	x6
high	relatively	moderate	normal	Fairly	warm	x7
high	relatively	moderate	normal	Partly	slightly cold	x8
moderate	relatively	moderate	moderate	Partly	slightly cold	x9
high	relatively	moderate	normal	Fairly	slightly cold	x10
						.
						.

الجدول ٤- ٢ نظام المعلومات الطقسي

كما ذكر سابقاً، تم الحصول على نظام المعلومات هذا بتطبيق خوارزمية الاستقراء الموجه بالصفة على قاعدة معطيات طقسية تتألف من٣٣ سجلاً. وتمثل الصفات في نظام المعلومات درجات الحرارة، وكميات السحب، والرطوبة النسبية. وسرع الرياح، والضغوط الجوية، وضغوط بحار الماء. وتوضح رموز الصفات والقيم الممكنة في الجدول ٤-٣.

Symbol	Attribute	Possible Values
a_1	Temperature	Very cold, Cold, Slightly cold, Warm, Moderate, Normal, Hot, Very Hot
a_2	Sky	Clear, Fairly, Partly, Cloudy
a_3	Humidity	Very low, Low, Relatively low, Normal, Moderate, Relatively high, High, Very high
a_4	Wind Speed	Calm, Light air, Light breeze, Moderate breeze, Strong breeze, Tornado, Hurricane
a_5	Air Pressure	Very low, Low, Relatively Low, Moderate, Relatively high, High, Very high
a_6	Vapor Pressure	Very low, low, moderate, high, Very high

الجدول ٤- ٣ رموز الصفات و قيمها المحتملة

٤-٤-٢ التجارب

في هذا المقطع الفرعي، نقدم نتائج بعض تجاربنا على نظام DBROUGH. ويكون الهدف الرئيسي لاختباراتنا إلقاء الضوء على السمات الرئيسية لنظامنا واختبار وظائفه. وأنجزت التجارب على معظم وظائف النظام. تمثل مجموعة المعطيات المستخدمة في تجاربنا بنظام معلومات الذي فيه الأشياء أو حالات الطقس في سياق مجالنا ترمز بالرمز x_i. وترمز الصفات بالرموز a1,a2,a3,a4,a5,a6 من أجل الصفات

درجة الحرارة، كمية السحب، الرطوبة النسبية، سرعة الرياح، الضغط الجوي، وضغط بخار الماء. والمجموعة C هي المجموعة الحاوية على الصفات الشرطية. والمجموعة D هي مجموعة صفات القرار، وعادة تحتوي على عنصر واحد، والمجموعة y هي مجموعة الأشياء التي تعرف مفهوم معين.

التجربة ١ : تحديد المفاهيم في نظام المعلومات الطقسي

يكون الهدف من هذه التجربة تحديد المفاهيم في نظام المعلومات لصفة قرار معينة بمعرفة المعلومات المتاحة.

بفرض أن صفة القرار هي a1، عندئذ المفاهيم في نظام المعلومات بالنسبة لهذه الصفة هي :

y1={x1, x2, x4, x5, x6, x8, x9, x10, x11, x12, x13, x14, x15, x17, x18, x19, x20},
y2={x3, x7, x16}.

بفرض أن صفة القرار هيa2 ، عندئذ المفاهيم في نظام المعلومات بالنسبة لهذه الصفة هي :

y1= {x1, x3, x8, x9, x15, x16, x17},
y2={x2, x4, x5, x6, x7, x10, x11, x12, x13, x14, x18, x19, x20}.

بفرض أن صفة القرار هيa3 ، عندئذ المفاهيم في نظام المعلومات بالنسبة لهذه الصفة هي :

y1= {x1, x2, x9, x13, x14, x15, x20},
y2= {x3, x4, x6, x7, x8, x10, x11, x16, x17, x18},
y3={x5, x12},
y4={x19}.

بفرض أن صفة القرار هيa4 ، عندئذ المفاهيم في نظام المعلومات بالنسبة لهذه الصفة هي :

y1= {x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x12},
y2= {x11},
y3= {x13, x14, x15, x16, x17, x18, x19, x20}.

بفرض أن صفة القرار هيa5 ، عندئذ المفاهيم في نظام المعلومات بالنسبة لهذه الصفة هي :

y1= {x1, x2, x3, x5, x6, x7, x8, x9, x10, x11, x12, x13, x16, x17},
y2= {x4, x14, x15, x18, x19, x20}.

بفرض أن صفة القرار هيa6 ، عندئذ المفاهيم في نظام المعلومات بالنسبة لهذه الصفة هي :

y1= {x1, x2, x3, x5, x7, x8, x10, x11, x13, x15, x16, x17, x18, x20},
y2= {x4, x6, x9, x12, x14, x19}.

التجربة ٢: استخراج السمات الملائمة

يكون الهدف من هذه التجربة تحديد المجموعة الفرعية الأساسية للصفات غير مكررة (غير زائدة)، المختزل الذي يعين مهام الاكتشاف لصفة قرار معينة. هذا يعني، استخراج السمات الملائمة من مجموعة معطاة من السمات. وفي هذه التجربة، سنقدم نتائج فحص السمات الملائمة لتعريف المفاهيم لصفات القرار المختلفة باستخدام الصفات الشرطية.

بفرض أن صفة القرار هي a1. عندئذ تكون السمات الملائمة لصفة القرار هذه بالنظر إلى كل المعلومات المتاحة، هذا يعني، بالنظر إلى المجموعة $C=\{a_2, a_3, a_4, a_5, a_6\}$ ، المجموعة $RED= \{a_3, a_4, a_6\}$. ويشير هذا أن الصفات a5 و a2 تكون زائدة أو غير ضرورية في المجموعة $C=\{a_2, a_3, a_4, a_5, a_6\}$ وبناء على ذلك، تكون مجموعة الصفات $RED= \{a_3, a_4, a_6\}$، التي تمثل الرطوبة وسرعة الرياح والضغط الجوي وضغط بخار الماء، فقط الصفات التي تكون قادرة على تمييز كل المفاهيم بالنسبة لصفة القرار a1 .

بفرض أن صفة القرار هي a2، عندئذ تكون السمات الملائمة لصفة القرار هذه بالنظر إلى كل المعلومات المتاحة، هذا يعني، بالنظر إلى المجموعة C={a1,a3,a4,a5,a6}، المجموعة RED={a3,a4,a5,a6}. ويشير هذا أن الصفات a3,a4,a5,a6 تكون ليست زائدة أو ضرورية في المجموعة C={a1,a3,a4,a5,a6}. وبناء على ذلك، تكون مجموعة الصفات RED={a3,a4,a5,a6}، التي تمثل الرطوبة، وسرعة الرياح، والضغط الجوي، وضغط بخار الماء، فقط الصفات التي تكون قادرة على تمييز كل المفاهيم بالنسبة لصفة القرار a2.

بفرض أن صفة القرار هي a3، عندئذ تكون السمات الملائمة لصفة القرار هذه بالنظر إلى كل المعلومات المتاحة، هذا يعني، بالنظر إلى المجموعة C={a1,a2,a4,a5,a6}، المجموعة RED={a1,a2,a4,a5,a6}. ويشير هذا أن الصفات a6,a5,a4,a2,a1 تكون ليست زائدة أو ضرورية في المجموعة C={a1,a2,a4,a5,a6}. وبناء على ذلك، تكون مجموعة الصفات RED={a1,a2,a4,a5,a6}، التي تمثل درج الحرارة، والسماء، وسرعة الرياح، والضغط الجوي، وضغط بخار الماء، فقط الصفات التي تكون قادرة على تمييز كل المفاهيم بالنسبة لصفة القرار a3.

بفرض أن صفة القرار هي،a4 عندئذ تكون السمات الملائمة لصفة القرار هذه بالنظر إلى كل المعلومات المتاحة، هذا يعني، بالنظر إلى المجموعة C={a1,a2,a3,a5,a6}، المجموعة $RED=\{ a_1, a2,a3,a5,a6 \}$ ويشير هذا أن الصفات a1,a2,a3,a5,a6 تكون ليست زائدة أو ضرورية في المجموعة C={a1,a2,a3,a5,a6} وبذلك، تكون مجموعة الصفات $RED=\{a_1,a2,a3,a5,a6\}$ التي تمثل درجة الحرارة والسماء، وسرعة الرياح، والضغط الجوي، وضغط بخار الماء، فقط الصفات التي تكون قادرة على تمييز كل المفاهيم بالنسبة لصفة القرار a4.

بفرض صفة القرار هي a5، عندئذ تكون السمات الملائمة لصفة القرار هذه بالنظر إلى كل المعلومات المتاحة، هذا يعني بالنظر إلى المجموعة. $C=\{a_1, a_2, a_3, a_4, a_6\}$ المجموعة $RED=\{a_2, a_3, a_4, a_6\}$. ويشير هذا أن الصفة a_1 تكون ليست زائدة أو ضرورية في المجموعة $C=\{a_1, a_2, a_3, a_4, a_6\}$. وبناءً على ذلك تكون مجموعة الصفات $RED=\{a_2, a_3, a_4, a_6\}$، التي تمثل السماء والرطوبة النسبية وسرعة الرياح وضغط بخار الماء، فقط الصفات التي تكون قادرة على تميز كل المفاهيم بالنسبة لصفة القرارa5.

بفرض أن صفة القرار هي a6، عندئذ تكون السمات الملائمة لصفة القرار هذه بالنظر إلى كل المعلومات المتاحة، هذا يعني، بالنظر إلى المجموعة $C=\{a_1, a_2, a_3, a_4, a_5\}$ المجموعة $RED=\{a_1, a_2, a_3, a_4, a_5\}$. ويشير هذا أن الصفات a_1, a_2, a_3, a_4, a_5 تكون غير زائدة أو ضرورية في المجموعة $C=\{a_1, a_2, a_3, a_4, a_5\}$. وبناءً على ذلك، تكون مجموعة الصفات $RED=\{a_1, a_2, a_3, a_4, a_5\}$. التي تمثل درجة الحرارة السماء والرطوبة النسبية وسرعة الرياح والضغط الجوي، فقط الصفات التي تكون قادرة على تميز كل المفاهيم بالنسبة لصفة القرار a6.

التجربة ٣ : توليد قواعد القرار الاحتمالية

يكون الهدف من هذه التجربة توليد قواعد القرار الاحتمالية لمفهوم معين بمعرفة المعلومات المتاحة من مجموعة معطيات الهدف. وفي هذه التجربة، سنقدم عينات من قواعد القرار الاحتمالية المولدة لكل المفاهيم لصفات القرار المختلفة.

بفرض أن صفة القرار هي a1. من التجربة ١، رأينا أن نظام المعلومات يملك مفهومين بالنسبة لهذه الصفة. كان إجمالي عدد القواعد التي ولدت ١٣. يوضح عدد القواعد المولدة للمفاهيم المختلفة في الجدول ٤-٤، حيث NC يرمز لعدد عينات التدريب المعطاة بواسطة القواعد.

Concepts	Number of Rules	NC
y1	11	17
y2	2	3

الجدول ٤- ٤ عدد القواعد المولدة للمفاهيم بالنسبة للصفة a1

تعرض عينة من قواعد القرار الاحتمالية المولدة لكل المفاهيم للصفةa1 في الشكل ٤-٤.

(a3=normal) $\wedge$ (a4=moderate breeze) $\wedge$ (a6=high) → (a1=slightly cold) (CF=0.57)
(a3=normal) $\wedge$ (a4=strong breeze) $\wedge$ (a6=high) → (a1=warm) (CF=0.67)

الشكل ٤- ٤ عينة من قواعد القرار الاحتمالية المولدة لكل المفاهيم للصفةa1

تنص القاعدة الأولى في الشكل ٤-٤ على أنه إذا كانت الرطوبة النسبية بين ٥٠ و٦٠ بالمئة وسرعة الرياح كانت بين ١٠و٢٢ عقدة وضغط بخار الماء كان بين ١٥ و٢٠ ميلي بار، عندئذ درجة الحرارة تكون بين ١٥ و١٠ درجة بمعامل تأكد (عامل مصداقية) ٠.٥٧.

تنص القاعدة الثانية في الشكل ٤-٤ على أنه إذا كانت الرطوبة النسبية بين٥٠ و٦٠ بالمئة وكانت سرعة الرياح بين ٢٢ و٣٠ عقدة وكان ضغط بخار الماء بين ١٥ و٢٠ ميلي بار، عندئذ درجة الحرارة تكون بين ١٥و٢٠ درجة بمعامل تأكد ٠.٦٧.

بفرض أن صفة القرار هي a2 من التجربة ١، رأينا أن نظام المعلومات يملك مفهومين بالنسبة لهذه الصفة. كان إجمالي عدد القواعد التي ولدت ١٦. يوضح عدد القواعد المولدة للمفاهيم المختلفة في الجدول ٤-٥، حيث NC يرمز إلى عدد عينات التدريب المغطاة بواسطة القواعد.

Concepts	Number of Rules	NC
y1	5	7
y2	11	13

الجدول ٤- ٥ عدد القواعد المولدة للمفاهيم بالنسبة للصفة a2

تعرض عينة من قواعد القرار الاحتمالية المولدة لكل المفاهيم للصفة a2 في الشكل ٤-٥.

(a3=moderate) $\wedge$ (a4=moderate breeze) $\wedge$ (a5=relatively high) $\wedge$ (a6=high) → (a2=partly) (CF=0.50)

(a3=relatively high) $\wedge$ (a4=strong breeze) $\wedge$ (a5=high)(a6=moderate) → (a2=fairly) (CF=1.00)

الشكل ٤- ٥ عينة من قواعد القرار الاحتمالية المولدة لكل المفاهيم للصفة a2

تنص القاعدة الأولى في الشكل ٤-٥ على أنه إذا كانت الرطوبة النسبية بين ٦٠ و٧٠ بالمئة وكانت سرعة الرياح بين ١٠ و٢٢ عقدة وكان الضغط الجوي بين ١٠١٥ و١٠٢٠ ميلي بار وكان ضغط بخار الماء بين ١٥ و٢٠ ميلي بار، عندئذ تكون كميات السحب بين ٢ و٥ أثمان بعامل تأكد ٠.٠٥.

تنص القاعدة الثانية في الشكل ٤-٥ على أنه إذا كانت الرطوبة النسبية بين ٧٠ و٨٠ بالمئة وكانت سرعة الرياح بين ٢٢ و٣٠ عقدة وكان الضغط الجوي بين ١٠٢٠ و١٠٢٥ ميلي بار وكان ضغط بخار الماء بين ١٠ و١٥ ميلي بار، عندئذ تكون كميات السحب بين ٠ و٢ ثمن بعامل تأكد ١.

بفرض أن صفة القرار هي a3 من التجربة ١ رأينا أن نظام المعلومات يملك ٤ مفاهيم بالنسبة لهذه الصفة. كان إجمالي عدد القواعد التي ولدت ٢٠. يوضح عدد القواعد المولدة للمفاهيم المختلفة في الجدول ٤-٦ ، حيث NC يرمز إلى عدد عينات التدريب المغطاة بواسطة القواعد.

Concepts	Number of Rules	NC
y1	7	7
y2	10	10
y3	3	2
y4	1	1

الجدول ٤– ٦ عدد القواعد المولدة للمفاهيم بالنسبة للصفة a3

تعرض عينة قواعد القرار الاحتمالية المولدة لكل المفاهيم للصفة a3 في الشكل ٤–٦.

(a1=slightly cold) ∧ (a2=partly) ∧ (a4=moderate breeze) ∧ (a5=relatively high)(a6=high) → (a3=moderate) (CF=0.60)
(a1=slightly cold) ∧ (a2=fairly) ∧ a4=moderate breeze) ∧ (a5=relatively high) ∧ (a6=high) → (a3=normal) (CF=0.67)
(a1=slightly cold) ∧ (a2=fairly) ∧ (a4=moderate breeze) ∧ (a5=relatively high) ∧ (a6=moderate) → (a3=relatively low) (CF=0.67)
(a1=slightly cold) ∧ (a2=fairly) ∧ (a4=strong breeze) ∧ (a5=high) ∧ (a6=moderate) → (a3=relatively high) (CF=0.75)

الشكل ٤- ٦ عينة من قواعد القرار الاحتمالية المولدة لكل المفاهيم للصفة a3

تنص القاعدة الأولى في الشكل ٤–٦ على أنه إذا كانت درجة الحرارة بين ١٠ و١٥ درجة وكانت كميات السحب بين ٢ و٥ أثمان وكانت سرعة الرياح بين ١٠ و٢٢ عقدة وكان الضغط الجوي بين ١٠١٥ و١٠٢٠ ميلي بار وكان ضغط بخار الماء بين ١٥ و٢٠ ميلي بار، عندئذ تكون الرطوبة النسبية بين ٦٠ و٧٠ بالمائة بعامل تأكد ٠.٦٠.

تنص القاعدة الثانية في الشكل ٤–٦ على أنه إذا كانت درجة الحرارة بين ١٠ و١٥ درجة وكانت كميات السحب بين ٠ و٢ ثمن وكانت سرعة الرياح بين ١٠ و٢٢ عقدة وكان الضغط الجوي بين ١٠١٥ و١٠٢٠ ميلي بار وكان ضغط بخار الماء بين ١٥ و٢٠ ميلي بار، عندئذ تكون الرطوبة النسبية بين ٥٠ و٦٠ بالمائة بعامل تأكد ٠.٦٧.

تنص القاعدة الثالثة في الشكل ٤–٦ على أنه إذا كانت درجة الحرارة بين ١٠ و١٥ درجة وكانت كميات السحب بين ٠ و٢ ثمن وكانت سرعة الرياح بين ١٠ و٢٢ عقدة وكان الضغط الجوي بين ١٠١٥ و١٠٢٠ ميلي بار وكان ضغط بخار الماء بين ١٠ و١٥ ميلي بار، عندئذ تكون الرطوبة النسبية بين ٧٠ و٨٠ بالمائة بعامل تأكد ٠.٦٧.

تنص القاعدة الرابعة في الشكل ٤–٦ على أنه إذا كانت درجة الحرارة بين ١٠ و١٥ درجة وكانت كميات السحب بين٠ و٢ ثمن وكانت سرعة الرياح بين ٢٢ و٣٠ عقدة وكان الضغط الجوي بين ١٠٢٠ و١٠٢٥ ميلي بار وكان ضغط بخار الماء بين ١٠ و١٥ميلي بار وكان ضغط بخار الماء بين ١٠ و١٥ ميلي بار، عندئذ تكون الرطوبة النسبية بين ٧٠ و٨٠ بالمائة بعامل تأكد ٠.٧٥.

بفرض أن صفة القرار هي a4 من التجربة ١. رأينا أن نظام المعلومات يملك ٣ مفاهيم بالنسبة لهذه الصفة. كان إجمالي عدد القواعد المولدة ٢٠. يوضح عدد القواعد المولدة للمفاهيم المختلفة في الجدول ٥-٦، حيث NC يرمز إلى عدد عينات التدريب المغطاة بواسطة القواعد.

Concepts	Number of Rules	NC
y1	11	11
y2	1	1
y3	8	8

الجدول ٤-٧ عدد القواعد المولدة للمفاهيم بالنسبة للصفة a4

تعرض عينة من قواعد القرار الاحتمالية المولدة لكل المفاهيم للصفة a4 في الشكل ٤-٧.

(a1=slightly cold) ∧ (a2=partly) ∧ (a3=moderate) ∧ (a5=relatively high) ∧ (a6=high) → (a4=moderate breeze) (CF=1.00)
(a1=slightly cold) ∧ (a2=fairly) ∧ (a3=normal) ∧ (a5=relatively high) ∧ (a6=high) → (a4=light breeze) (CF=0.67)
(a1=slightly cold) ∧ (a2=fairly) ∧ (a3=moderate) ∧ (a5=relatively high) ∧ (a6=high) → (a4=strong breeze) (CF=0.50)

الشكل ٤-٧ عينة من قواعد القرار الاحتمالية المولدة لكل المفاهيم للصفة a4

تنص القاعدة الأولى في الشكل ٤-٧ على أنه إذا كانت درجة الحرارة بين ١٠ و١٥ درجة وكانت كميات السحب بين ٢ و٥ أثمان وكانت الرطوبة النسبية بين ٦٠ و٧٠ بالمئة وكان الضغط الجوي بين ١٠١٥ و١٠٢٠ ميلي بار وكان ضغط بخار الماء بين ١٥ و٢٠ميلي بار، عندئذ تكون سرعة الرياح بين ١٠ و٢٠ عقدة بعامل تأكد ١.

تنص القاعدة الثانية في الشكل ٤-٧ على أنه إذا كانت درجة الحرارة بين ١٠ و١٥ درجة وكانت كميات السحب بين ٠ و٢ ثمن وكانت الرطوبة النسبية بين ٥٠ و٦٠ بالمئة وكان الضغط الجوي بين ١٠١٥ و١٠٢٠ ميلي بار وكان ضغط بخار الماء بين ١٥ و٢٠ ميلي بار، عندئذ تكون سرعة الرياح بين ٤ و١٠ عقدة بعامل تأكد ٠.٦٧.

تنص القاعدة الثالثة في الشكل ٤-٧ على انه إذا كانت درجة الحرارة بين ٢٠ و٢٥ درجة وكانت كميات السحب بين ٠ و٢ ثمن وكانت الرطوبة النسبية بين ٦٠ و٧٠بالمئة وكان الضغط الجوي بين ١٠١٥ و١٠٢٠ ميلي بار وكان ضغط بخار الماء بين ١٥ و٢٠ ميلي بار، عندئذ تكون سرعة الرياح بين ٢٢ و٣٠ عقدة بعامل تأكد ٠.٥٠.

بفرض أن صفة القرار هي a5 من التجربة ١. رأينا أن نظام المعلومات يملك مفهومان بالنسبة لهذه الصفة. كان إجمالي عدد القواعد المولدة ١٧. يوضح عدد القواعد المولدة للمفاهيم المختلفة في الجدول ٤-٨، حيث NC يرمز لعدد عينات التدريب المغطاة بواسطة القواعد.

Concepts	Number of Rules	NC
y1	11	14
y2	6	6

الجدول ٤-٨ عدد القواعد المولدة للمفاهيم بالنسبة للصفة a5

تعرض عينة عن قواعد القرار الاحتمالية المولدة لكل المفاهيم للصفة a5 في الشكل ٤-٨.

(a2=partly) ∧ (a3=moderate) ∧ (a4=moderate breeze) ∧ (a6=high) → (a5=relatively high) (CF=1.00)

(a2=fairly) ∧ (a3=moderate) ∧ (a4=strong breeze) ∧ (a6=high) → (a5=high) (CF=0.75)

الشكل ٤-٨ عينة من قواعد القرار الاحتمالية المولدة لكل المفاهيم للصفة a5

تنص القاعدة الأولى في الشكل ٤-٨ على أنه إذا كانت كميات السحب بين ٢و٥ أثمان وكانت الرطوبة النسبية بين ٦٠ و٧٠ بالمئة وكانت سرعة الرياح بين١٠ و٢٢ عقدة وكان ضغط بخار الماء بين ١٥ و٢٠ ميلي بار، عندئذ يكون الضغط الجوي بين ١٠١٥ و١٠٢٠ ميلي بار بعامل تأكد ١.

تنص القاعدة الثانية في الشكل ٤-٨ على أنه إذا كانت كميات السحب بين ٠ و٢ ثمن وكانت الرطوبة النسبية بين ٦٠ و٧٠ بالمئة وكانت سرعة الرياح بين ٢٢ و٣٠ عقدة وكان ضغط بخار الماء بين ١٥ و٢٠ ميلي بار، عندئذ يكون الضغط الجوي بين ١٠٢٠ و١٠٢٥ ميلي بار بعامل تأكد ٠.٧٥.

بفرض انه صفة القرار هي a6 من التجربة ١. رأينا أن نظام المعلومات يملك مفهومان بالنسبة لهذه الصفة. كان إجمالي عدد القواعد المولدة ٢٠ قاعدة. يوضح عدد القواعد المولدة للمفاهيم المختلفة في الجدول ٤-٩، حيث NC يرمز إلى عدد عينات التدريب المغطاة بواسطة القواعد.

Concepts	Number of Rules	NC
y1	14	14
y2	6	6

الجدول ٤-٩ عدد القواعد المولدة للمفاهيم بالنسبة للصفة a6

تعرض عينة من قواعد القرار الاحتمالية المولدة لكل المفاهيم للصفة a6 في الشكل ٤-٩.

(a1=slightly cold) $\wedge$ (a2=partly) $\wedge$ (a3=moderate) $\wedge$ (a4=moderate breeze) $\wedge$ (a5=relatively high)->(a6=high) (CF=0.75)
(a1=slightly cold) $\wedge$ (a2=fairly)(a3=relatively low) $\wedge$ (a4=moderate breeze) $\wedge$ (a5=relatively high)-> (a6=moderate) (CF=0.50)

الشكل ٤- ٩ عينة من قواعد القرار الاحتمالية المولدة لكل المفاهيم للصفة a6

تنص القاعدة الأولى في الشكل ٤-٩ على أنه إذا كانت درجة الحرارة بين ١٠ و١٥ درجة وكانت كميات السحب بين ٢ و٥ أثمان وكانت الرطوبة النسبية بين ٦٠ و٧٠ بالمئة وكانت سرعة الرياح بين ١ و٢٢ عقدة وكان الضغط الجوي بين ١٠١٥ و١٠٢٠ ميلي بار، عندئذ يكون ضغط بخار الماء بين ١٥ و٢٠ ميلي بار بعامل تأكد ٠.٧٥.

٤-٥ تقييم الخبراء للقواعد المكتشفة

قبلت القواعد المولدة من قبل خبراء المجال. وكانت معظمها ذات جودة عالية وكانت قادرة على تحديد ما يعتقده الخبير علاقات منطقية في مجال الدراسة. علاوة على ذلك، كانت أغلبها قواعد ممكنة بعامل تأكد مقبول إلى حد ما. وهذا يعني أنها تعرف المفاهيم التي تكون غامضة إلى حد ما. بالإضافة إلى ذلك، إنها تعكس بعض الخواص الحقيقية في المعطيات.

٤-٦ ملاحظات ختامية

في هذا الفصل، اقترحنا طريقة لتوليد قواعد التنبؤ بالطقس من مجموعة من أمثلة التدريب. وفي الواقع بنيت الطريقة على (آ) طريقة الاستقراء الموجة بالصفة بعد دراسة تأثير عناصر الطقس على بعضها البعض باستخدام نموذج الانحدار المتعدد و(ب) تمديد نموذج المجموعات التقريبية. تكون السمة البارزة الأولى للطريقة المقترحة أنها تأخذ بعين الاعتبار تأثير العناصر على بعضها البعض وتكون السمة الثانية للطريقة المقترحة أنها تستفيد من المعلومات الإحصائية في نظام المعرفة. وهكذا، تكون طريقتنا قادرة على اشتقاق قواعد القرار من المعرفة غير كاملة. وتكون هذه المقدرة هامة نظراً لأننا نادراً ما نملك معلومات كاملة ومتسقة في تصميم النظم التنبؤية. وتكون السمة الثالثة للطريقة المقترحة أننا غيرنا المجال [0. 5,1) للعددβ المقترح (المفروض) من قبل ايجان أن وآخرون إلى المجال [0. 25,1) لأنه عندما كانت

β عدد حقيقي في المجال[0. 5,1)، كل الصفات الشرطية لنظام المعلومات الموصوف في الجدول ٤ .١ أصبحت غير ضرورية. وهذا ليس صحيحا لأن الصفات a2 و a1 هي صفات ضرورية.

ويملك تطبيق طريقة اكتشاف المعرفة للتنبؤ بالطقس ميزة أن علاقات المصاحبة الهام بين الصفات الشرطية وصفة القرار توصف بلغة قواعد "IF-THEN" البسيطة، التي يمكن أن تفهم بسهولة من قبل المستخدمين.

و بنينا نظام تعلم قاعدة معطيات تجريبي، DBROUGH، لاكتشاف قواعد المعرفة من قواعد المعطيات العلاقية ضمن إطار عمل منهج نظرية المجموعات التقريبية ونموذج تحليل الانحدار. ومن ثم أجرينا بعض التجارب الهامة في نظام التعلم هذا في تطبيق مناخي. وسردنا بعض النتائج التجريبية للتطبيق. وبعد ذلك عرضنا تقييم لقواعد المعرفة المكتشفة من قبل خبير أرصاد جوية.

كان الهدف الرئيسي للتطبيق اكتشاف قواعد المعرفة من المعطيات للتنبؤ الشهري بدرجات الحرارة وكميات السحب والرطوبة النسبية وسرع الرياح والضغوط الجوية وضغوط بخار الماء. وستستخدم هذه القواعد المكتشفة فيما بعد في بناء نظام خبير للتنبؤ.

يمكن أن تلخص نتائج كل التجارب كما يلي :

١. كما نوقش نادراً ما تستخدم قيم الصفات الكمية مباشرةً في تحليل المجموعات التقريبية. عوضاً عن ذلك، يقسم المجال الأصلي لكل صفة إلى بضعة فترات فرعية تمثل المصطلحات الكمية. في تطبيقنا، عمل فصل للمعطيات وفقاً لمجال الدراسة وقيم من قبل خبير المجال (النطاق).
٢. يستقرئ نظام التعلم DBROUGH نظام المعلومات بتطبيق خوارزمية الاستقراء الموجه بالصفة.
٣. يستطيع نظام التعلم DBROUGH أن يحدد المفاهيم في نظام المعلومات لصفة قرار معينة بمعرفة المعلومات المتاحة.
٤. يستطيع نظام التعلم DBROUGH أن يحدد المجموعة الفرعية الأساسية للصفات غير مكررة (غير زائدة) التي تعين مهمة اكتشاف لصفة قرار معينة. هذا يعني، يستطيع النظام أن يستخرج السمات الملائمة من مجموعة محددة من السمات.
٥. يستطيع نظام التعلم DBROUGH أن يولد قواعد المعرفة مع عوامل التأكد لمفهوم معين بمعرفة المعلومات المتاحة من مجموعة معطيات الهدف.
٦. تحدد كل قواعد المعرفة المكتشفة ما يعتقده الخبير كعلاقة منطقية في المجال وهكذا قبلت هذه القواعد من قبل الخبير.

الفصل الخامس

تنفيذ نظام خبير استقرائي للتنبؤ بالطقس

٥-١ مقدمة

يقدم هذا الفصل التنفيذ للنظام الخبير الاستقرائي الفريد للتنبؤ بالطقس، IESP. ويكون نظـام IESP نظـام مبني على الاستقراء يوظف نموذج تحليل الانحدار ونظرية المجموعات التقريبية لاكتشاف أنمـاط الطقـس العامة والقواعد من قواعد المعطيات الطقسية. وينظم هذا الفصل كما يلي :

يقدم المقطع ٥-٢ تنفيذ النظام المقترح. ويقدم المقطع ٥-٣ السمات الرئيسية لنظام IESP.وفـي المقطـع ٥-٤ تم شرح تشغيل النظام، وأخيراً، يختتم الفصل. بملاحظات ختامية في المقطع ٥-٥.

٥-٢ تنفيذ النظام

توجد عدة طرق لبناء النظم الخبيرة. تكـون إحـدى الطـرق باسـتخدام الأغلفـة مثـل VP-ExPERT وKAPPA-PC وExSyS وROSIE وN-ExPERT وCLIPS وLEVEL5 إلـخ. وتكـون الطريقـة الأخرى بالبرمجة باستخدام LISP وPROLOG، أو باستخدام بعض لغات البرمجة المناسبة الأخرى مثـل الفيجوال ++C و#C والدلفي. ومن أجل مشكلتنا، لقد تم اختيار #C من أجل برمجة هذا النظام وهي لغـة تدعم مفهوم البرمجة غرضية التوجه object-oriented programming (OOP). حيـث يعطـي هـذا المفهوم أسلوب أكثر طبيعية في التعامل مع عملية البرمجة وذلك من خلال نمذجته لأغراض العالم الحقيقي عن طريق تطوير بنى مرنة (نظم وبرامج) مماثلة لها، فهو يستفيد من ميزات العلاقات بين الأغراض التـي تنتمي إلى وصف معين، كما أنه يستفيد من علاقات التوريث، حيث يجري ذلك بأن ترث الصفوف صـفات الصفوف الموجودة مسبقاً بالإضافة إلى الصفات الخاصة التي تتمتع بها الأغراض التي نهتم بها. ومن أهـم الميزات التي اعتمدنا عليها في اختيار لغة البرمجة #C في تنفيذ نظامنا المقترح:

١- لقد تخلصت # **C** من التعقيدات والمشاكل الخاصة باللغات مثل java و++C فقد قامت بإلغاء الماكرو والقوالب والتوارث المتعدد وطبقات الأساس الظاهرية فهذه هي المناطق التي تظهر فيها الالتباسات لدى المبرمجين وكذلك ظهور المشاكل المعقدة.

٢- تحتوي #C على أدوات حديثه لم تكن موجودة في أي لغة من قبل ألا وهي السمات وهي طريقة قام مصممي الـ # Cبتضمينها لم توجد في أي لغة أخرى والسمات هي عبارة عن صفوف لها خاصية ربط المعلومات على أي أوامر بـ # Cويمكن الحصول على هذه المعلومات أثناء التشغيل وتتوارث السمات من صف يدعى Attribute هذا يجعلها لها خاصية ربط المعلومات وتوجد

خصائص أخر للـــ C # مثل معالجتها للاستثناءات العالية الكفاءة. كما أن C # لا تسمح بوقوع الأخطاء السخيفة.

٣- C # لغة برمجية تستخدم الكائنات، من الواضح تفوق البرمجة باستخدام الكائنات على البرمجة المهيكلة كثيراً بما تسببه البرمجة المهيكلة مثل برمجة C صعوبة في اكتشاف الأخطاء وصعوبة في الصيانة لأنها تكتب مرة واحدة كتلة من التعليمات.

٤- الكلمات الأساسية قليلة: تمتاز لغة **C #** بعدد قليل من الكلمات وهذا لا يدل على أنها ضعيفة بل هذه احد مزايا اللغة.

٥- الاقتصاد في الذاكرة (عملية جمع القمامة): أهم مميزات **C #** التزويد بآلية جبارة وآمنة لجمع القمامة في وقت التشغيل أيضاً مما يضفي على مبرمجي **C #** مزيد من التحكم بالكائنات، خلافاً لما تفعله لغة **C++** بوضع كاهل التخلص من الكائنات على عاتق المبرمج. حيث تعتبر مشكلة تسرب الذاكرة من أهم المشاكل التي نصادفها في لغات أخرى، إذ يحتل كل غرض عند إنشائه حيزاً من الذاكرة، وتبقى هذه الذاكرة مشغولة حتى بعد انتفاء الحاجة للغرض، وتقع مسؤولية تحرير مواقع الأغراض غير المستخدمة على عاتق المبرمج، أما في **C #** فتقع مسؤولية تحرير مواقع الذاكرة التي تشغلها أغراض غير مستخدمة على عاتق جامع النفايات الذي يقوم بالبحث عن تلك المواقع وتحريرها فور انخفاض مستوى الذاكرة الحرة عن حد معين. (تمتلك java آلية لجمع القمامة ولكنها أقل كفاءة من التي تستخدمها **C #**).

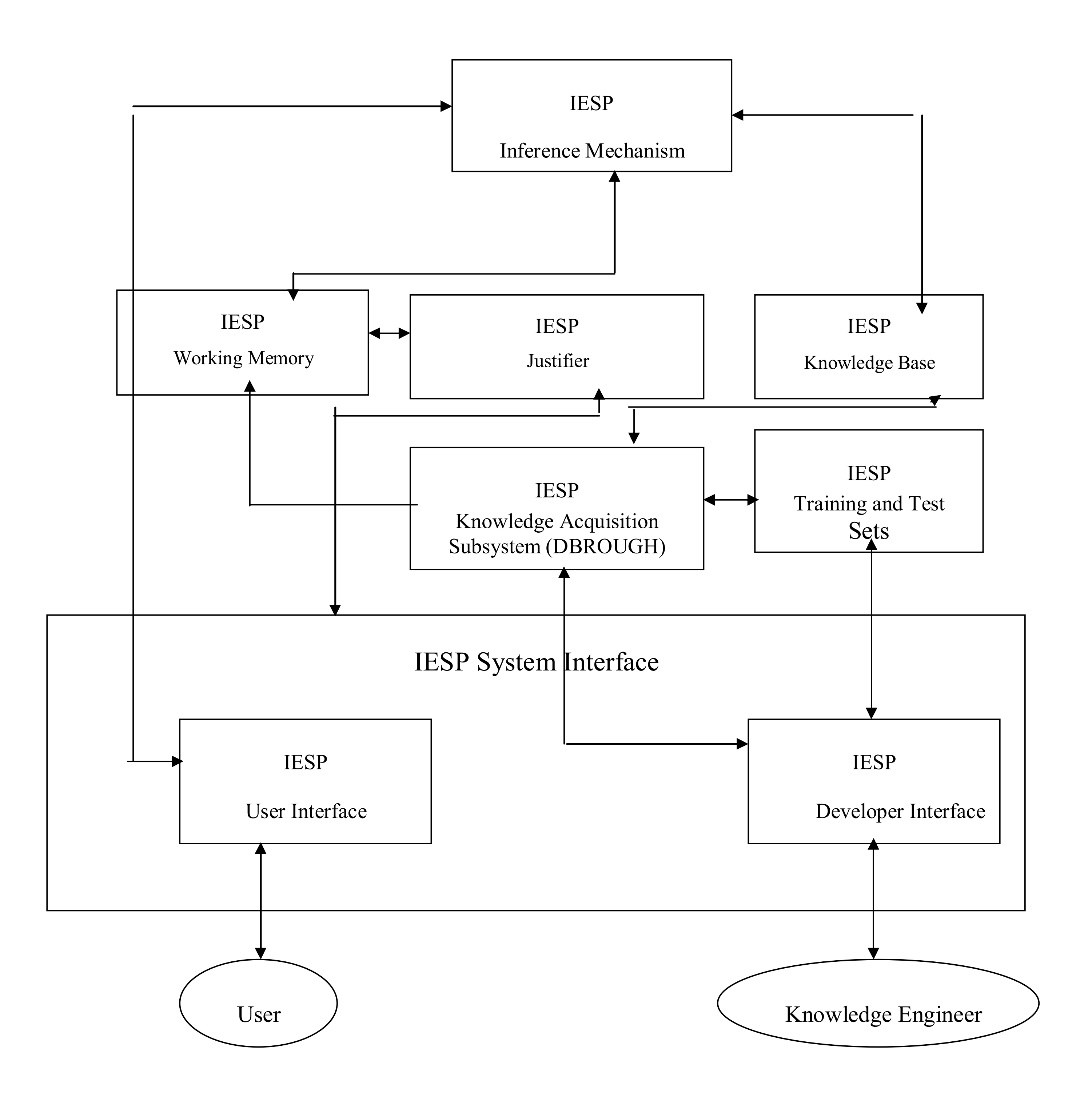
٥-٣ السمات الرئيسية لنظام IESP

في هذا المقطع، عرضت السمات الرئيسية لنظام IESP المنفذ. أولاً، تم وصف معمارية نظام IESP. ثانياً، عرضت التوصيفات الكاملة للمكونات المختلفة المؤلفة للنظام.

٥-٣-١ معمارية نظام IESP

يكون نظام IESP نظام خبير استقرائي للتنبؤ بالطقس، يستغل نموذج تحليل الانحدار ومنهج نظرية المجموعات التقريبية الموجه بالصفة لاكتشاف أنماط الطقس العامة والقواعد من قواعد المعطيات الطقسية (مجموعات التدريب والاختبار).

وتوضح المعمارية لنظام IESP المقترح في الشكل ٥-١. كما نرى من هذا الشكل، يتألف نظام IESP المقترح من سبعة مكونات رئيسية: السطح البيني لنظام IESP، مجموعات التدريب والاختبار، ونظام فرعي لاكتساب المعرفة، وقاعدة معرفة، والذاكرة المستخدمة، وآلية الاستدلال (محرك الاستدلال)، ونظام التوضيح الفرعي (المبرر). وستشرح هذه المكونات فيما بعد.



الشكل ٥-١ معمارية النظام IESP

٥-٣-٢ وصف النظام

كما ذكر سابقاً، يتألف نظام IESP من سبعة مكونات رئيسية. في هذا المقطع الفرعي، نصف بصورة مختصرة كل مكونة من هذه المكونات السبعة.

٥-٣-٢-١ السطح البيني لنظام IESP

تقدم برمجيات تطوير النظم الخبيرة، بما فيها كلاً من الأغلفة واللغات الأساسية، أوجه مختلفة لكلاً من المستخدم ومطور النظام [١٢].

يكون السطح البيني لنظام IESP مكونة رئيسية في النظام المنفذ والتي تكون مسؤولة عن استخدام النظام. و السطح البيني لنظام IESP هو سطح بيني رسومي للمستخدم (GUI). ويعكس تصميم السطح البيني للنظام محتويات نظام IESP. ويعتبر السطح البيني لنظام IESP مؤلفاً من كل من السطح البيني للمطور والسطح البيني للمستخدم. ويوضح السطح البيني للنظام في الشكل ٥-٢.

ASTRO Login Interface

Using Inductive Expert Systems in Prediction
Using
Attribute_Oriented Rough Set Theory

IESP System Vresion 1.0

If You would likt to inter to forecasting system write "user" else If You would like to enter the DBROUGH,Training and test sets write "expert"

User Name

Password

Login Exit

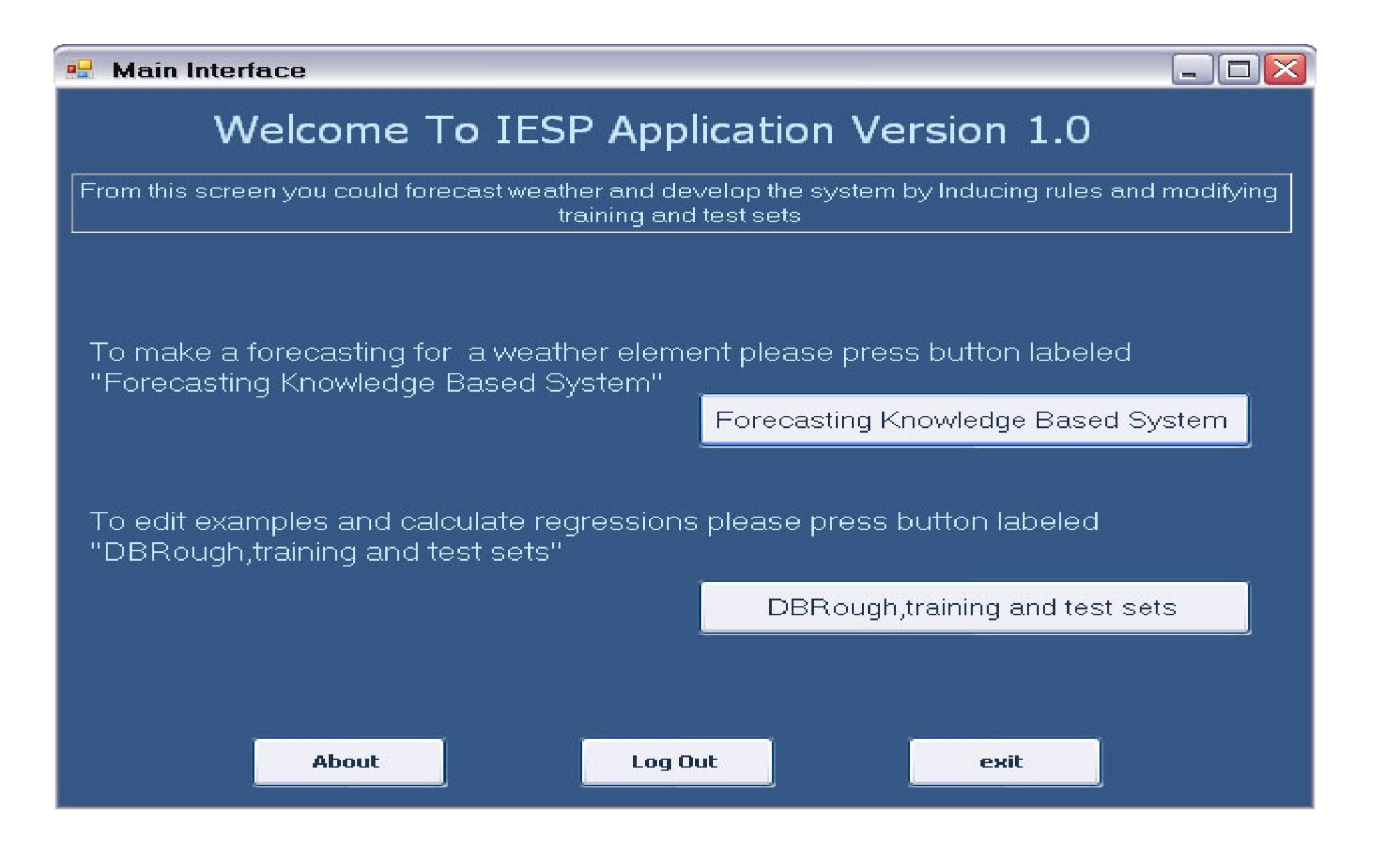
الشكل ٥-٢ السطح البيني للنظام IESP

كما نرى تحتوي هذه الشاشة على خيارات الدخول إلى النظام، فمن أجل عرض السطح البيني للمستخدم يجب على المستخدم الدخول كـ "user" وبالتالي تظهر الشاشة "User Interface". أما من

أجل عرض السطح البينـــي للمطـــور يجـــب علـــى المســـتخدم الـــدخول كـــــ "expert" حيـــث تظهـــر الشاشة"Developer Interface". ويكون الزر "Exit " مسؤولاً عن الخروج من النظام.

السطح البيني للمطور IESP

يكون السطح البيني للمطور IESP، الأداة التي من خلالها يطور مهندس المعرفة النظام. بكلمات أخـــرى، إنه الأداة التي من خلالها يطور مهندس المعرفة مجموعات التدريب والاختبار ويستقرئ القواعد لتطـــوير المعرفة للنظام. ويوضح السطح البيني للمطور IESP في الشكل ٥ –٣.

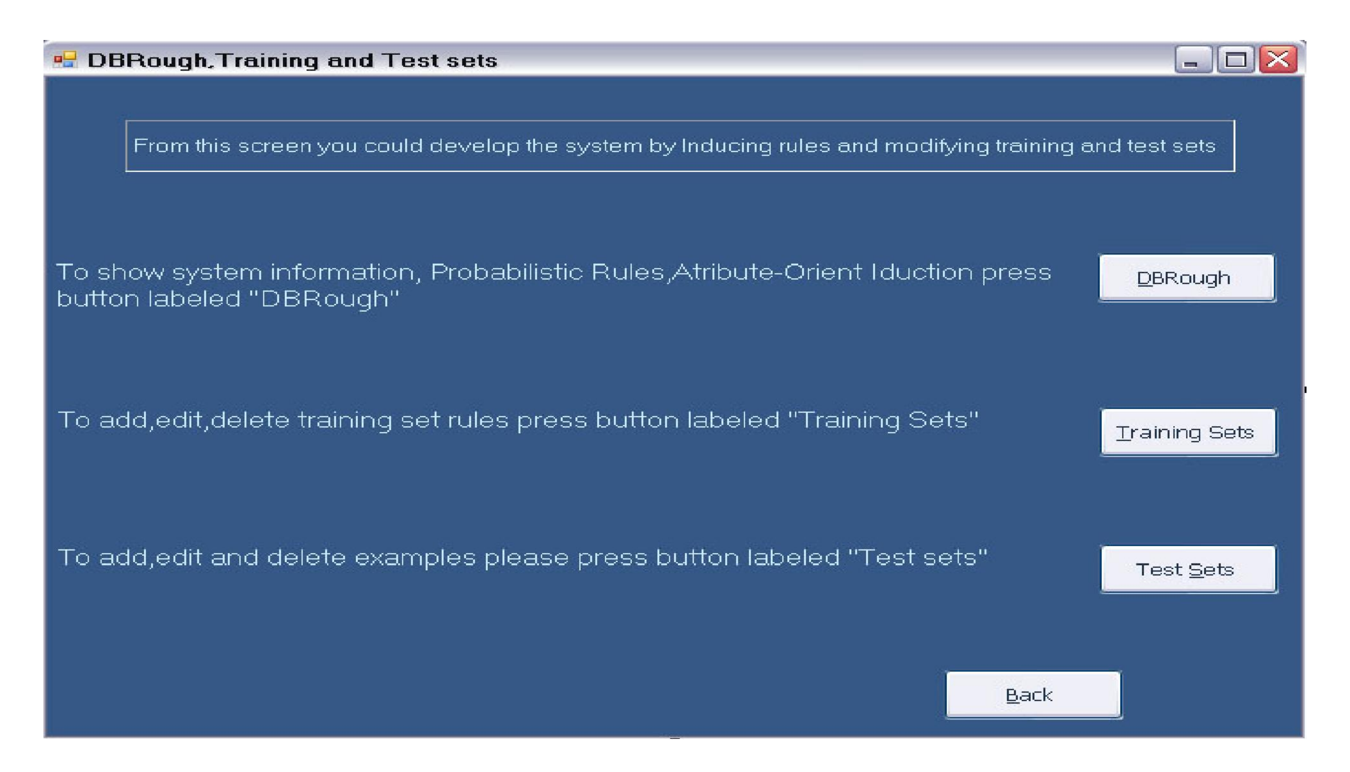


الشكل ٥–٣ السطح البيني للمطور

كما نرى تحتوي هذه الشاشة على خمسة أزرار، الزر الأول "DBRough,Training and Test Sets" يكون مسؤولاً عـــن عـــرض الشاشـــة "DBRough,Training and Test Sets"، أمـــا الـــزر "Forcasting Knowledge Based System" يكون مسؤول عن الدخول إلى النظام للتنبؤ لأن الخبيـــر يمكنه الدخول إلى النظام من أجل التنبؤ، كما يمكنه التطوير في مجموعات التدريب. ويكـــون الـــزر "log Out" مسؤولاً عن العودة إلى شاشة الدخول الرئيسية. ويكون الزر "about" مسؤولاً عن عرض الشاشـــة

المسؤولة عن عرض معلومات حول برمجة النظام. ويكون الزر "Exit" مسؤولاً عن الخروج من السطح البيني للمطور.

الشاشة "DBRough,Training and Test Sets"

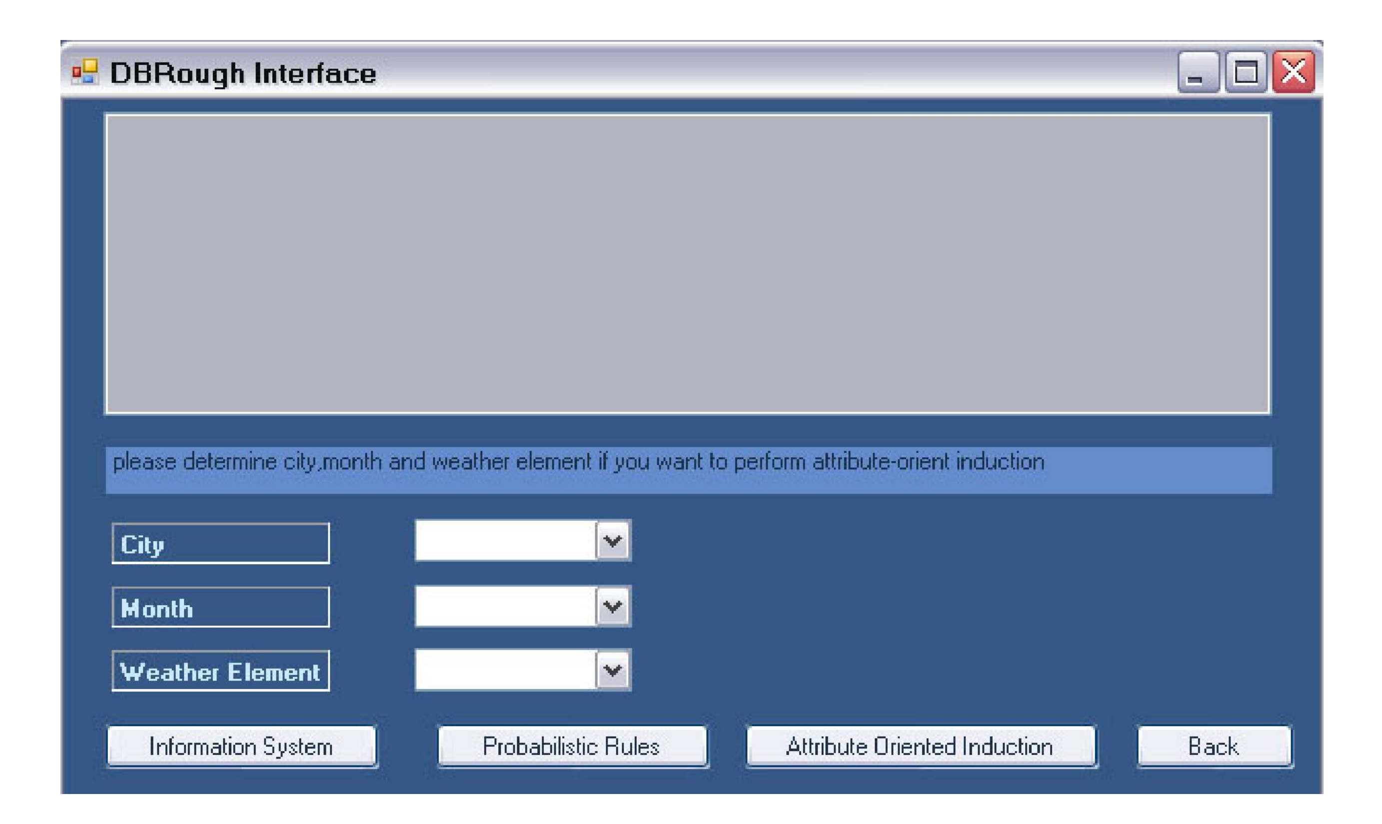


الشكل ٥–٤ شاشة DBRough,Training and Test Sets

تحتوي هذه الشاشة على أربعة أزرار "DBRough" و"Training sets" و"Test Sets" و"Back". يكون الزر "DBrough" مسؤولاً عن عرض الشاشة "DBRough Interface"، ويكون الزر "Training sets" مسؤولاً عن عرض الشاشة " Training sets". ويكون الزر " Test Sets " مسؤولاً عن عرض الشاشة " Test Sets Interface". ويكون الزر "Back" مسؤولاً عن العودة إلى السطح البيني للمطور.

(أ) شاشة "DBRough Interface"

الشاشة "DBRough Interface" هي الأداة التي من خلالها يستقرئ مهندس المعرفة نظام المعلومات وقواعد القرار الاحتمالية لتطوير قاعدة المعرفة لنظام IESP. وتوضح هذه الشاشة في الشكل ٥ – ٥.

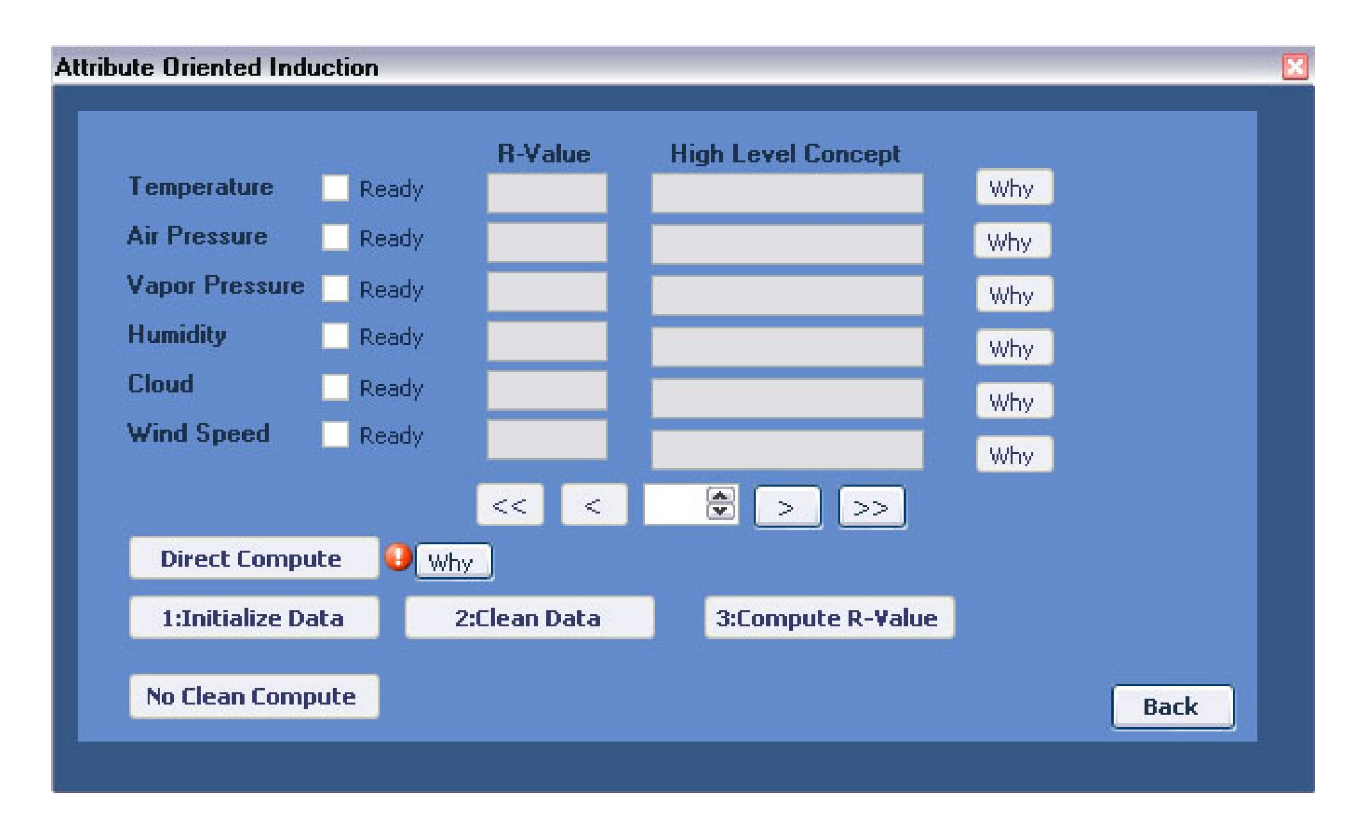


الشكل ٥-٥ شاشة DBRough Interface

تحتوي هذه الشاشة على أربعة أزرار، حيث يكون الزر "Information System" مسؤولا عن عرض نظام المعلومات لعنصر الطقس المختار، و يقوم الزر "Probabilistic Rules" بعرض قواعد القرار الاحتمالية لعنصر الطقس المختار، أما الزر "Attribute Oriented Induction " فهو مسؤول عن عرض الشاشة "Attribute Oriented Induction " التي تظهر في الشكل ٥ -٦، ويكون الزر "Back" مسؤولاً عن العودة إلى الشاشة "DBRough,Training and Test Sets".

الشاشة "Attribute Oriented Induction"

الشاشة "Attribute Oriented Induction" هي الأداة التي من خلالها يمكن لمهندس المعرفة أن يحسب قيمة الانحدار لمجموعات التدريب لعناصر الطقس المدخلة من قبله بالإضافة إلى ظهور المفاهيم العالية المستوى لهذه القيم.

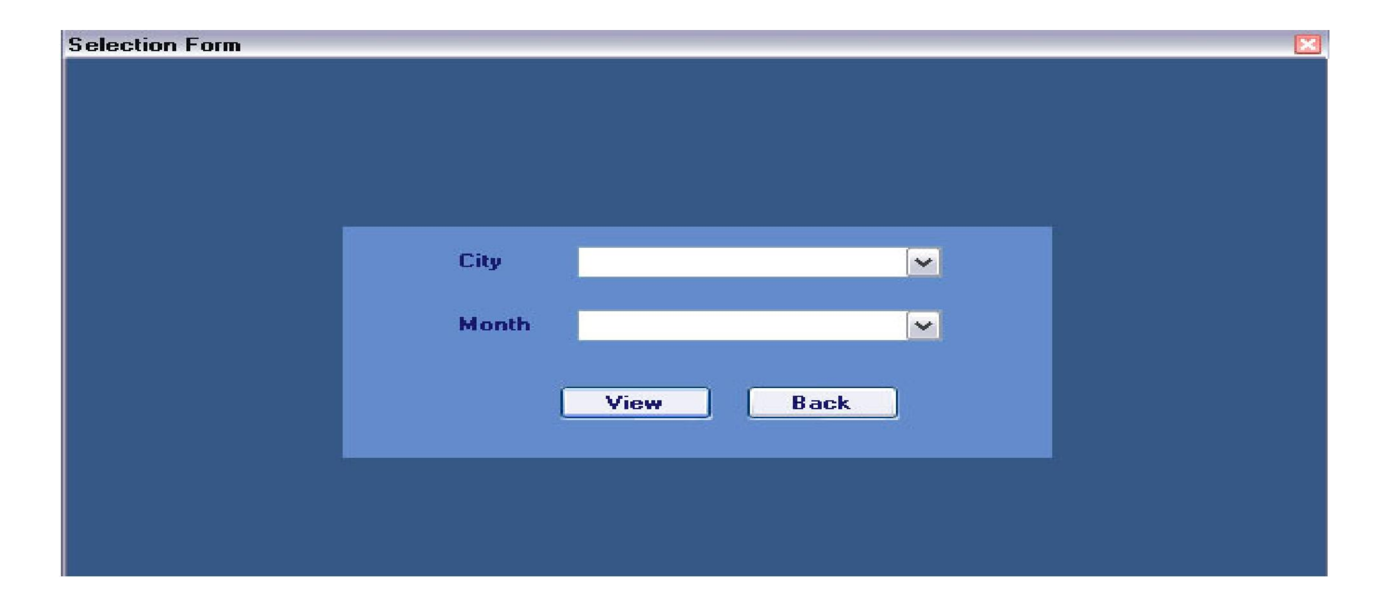


الشكل ٥-٦ شاشة Attribute Oriented Induction

يظهر في هذه الشاشة التلميح "ready" الذي يشير إلى عنصر الطقس الجاهز وعنصر الطقس الذي لم يتم إدخال قيم فيه، وذلك نظراً لأن حساب قيمة الانحدار تتم بعد إدخال قيم كافة عناصر الطقس، كما تظهر في هذه الشاشة خمس دوال أساسية، الدالة "Direct Compute" تقوم بحساب الانحدار مباشرة بعد إجراء عملية البحث عن قيم شاذة وإجراء عملية تنظيف البيانات، والدالة "Intialize data" تكون مسؤولة عن الكشف عن وجود قيم شاذة في القيم المدخلة لعناصر الطقس المحددة ففي حال وجودها تظهر إشارة تنبيه حمراء بجانب عنصر الطقس، أما الدالة "Clean Data" فتقوم بتنظيف القيم المدخلة من القيم الشاذة في حال وجودها، بينما الدالة "Compute R-value " فتقوم بحساب قيمة الانحدار بعد عملية تنظيف القيم من القيم الشاذة، وأخيراً الدالة "No Clean Compute" تكون مسؤولة عن حساب الانحدار للقيم المدخلة لعناصر الطقس بدون إجراء تنظيف للبيانات. وتوضح هذه الشاشة في الشكل ٥-٦.

(ب) شاشة Training Set

تظهر أولاً شاشة الاختيار **"Selection Form"** للحصول على معلومات من الخبير حول المدينة والشهر، الموضحة في الشكل ٥-٧.



الشكل ٥-٧ شاشة Selection Form

و تحتوي هذه الشاشة على الزر"view" ويكون مسؤول عن ظهـــور الشاشـــة "Training Set" للمدينة والشهر المدخلين كما في الشكل ٥ – ٨. والزر "Back" ويكون مسؤول عن العودة إلـــى الشاشـــة **"DBRough,Training and Test Sets"**.

الشكل ٥-٨ شاشة Training Set

والشاشة "Training Set" هي الأداة التي من خلالها يطور مهندس المعرفة الأمثلة لمجموعة التدريب بخصوص درجات الحرارة وكميات السحب والرطوبة النسبية وسرعة الرياح والضغط الجوي وضغط بخار الماء لنظام IESP. وتوضح هذه الشاشة في الشكل ٥-٨.

(ت) شاشة Element Test Set Interface

شاشة Element Test Set Interface هي الأداة التي من خلالها يطور مهندس المعرفة الأمثلة لمجموعة اختبار عناصر نظام IESP. حيث تظهر أولا الشاشة التي يمكن من خلالها اختيار العنصر المراد تطوير الأمثلة حوله كما في الشكل ٥-٩.

الشكل ٥-٩ شاشة Element Test Set Interface

وتحتوي هذه الشاشة على سبعة أزرار، فالزر "Air Pressure" يكون مسؤولا عن ظهور الشاشة "Air Pressure"، وهي الأداة التي من خلالها يطور مهندس المعرفة أمثلة مجموعة اختبار الضغط الجوي للنظام IESP، كما يظهر في الشكل ٥ – ١٠. أما الزر "Temperature" يكون مسؤولا عن ظهور الشاشة "Temperature "، وهي الأداة التي من خلالها يطور مهندس المعرفة أمثلة مجموعة اختبار درجة الحرارة للنظام IESP. ويكون الزر "Cloud" مسؤولا عن ظهور الشاشة "Cloud "، وهي الأداة التي من خلالها يطور مهندس المعرفة أمثلة مجموعة اختبار السحب للنظام IESP. بينما الزر

"Humidity" يكون مسؤولا عن ظهور الشاشة "Humidity "، وهي الأداة التي من خلالها يطور مهندس المعرفة أمثلة مجموعة اختبار الرطوبة للنظام IESP. أما الزر "Vapor Pressure" يكون مسؤولا عن ظهور الشاشة "Vapor Pressure "، وهي الأداة التي من خلالها يطور مهندس المعرفة أمثلة مجموعة اختبار ضغط بخار الماء للنظام IESP. والزر "Wind Speed" يكون مسؤولا عن ظهور الشاشة "Wind Speed "، وهي الأداة التي من خلالها يطور مهندس المعرفة أمثلة مجموعة اختبار سرعة الرياح للنظام IESP. وذلك بعد أن تظهر شاشة الاختيار **"Selection Form"** كما في الشكل ٥-٧ لاختيار المدينة و الشهر.

air_pressure

Year 1
Year 2
Year 3
Year 4
Year 5
Year 6
Year 7
Year 8
Year 9
Year 10
R-Value
Concept
<< < 1 > >>
Save Update Delete Back

الشكل ٥-١٠ شاشة Air Pressure

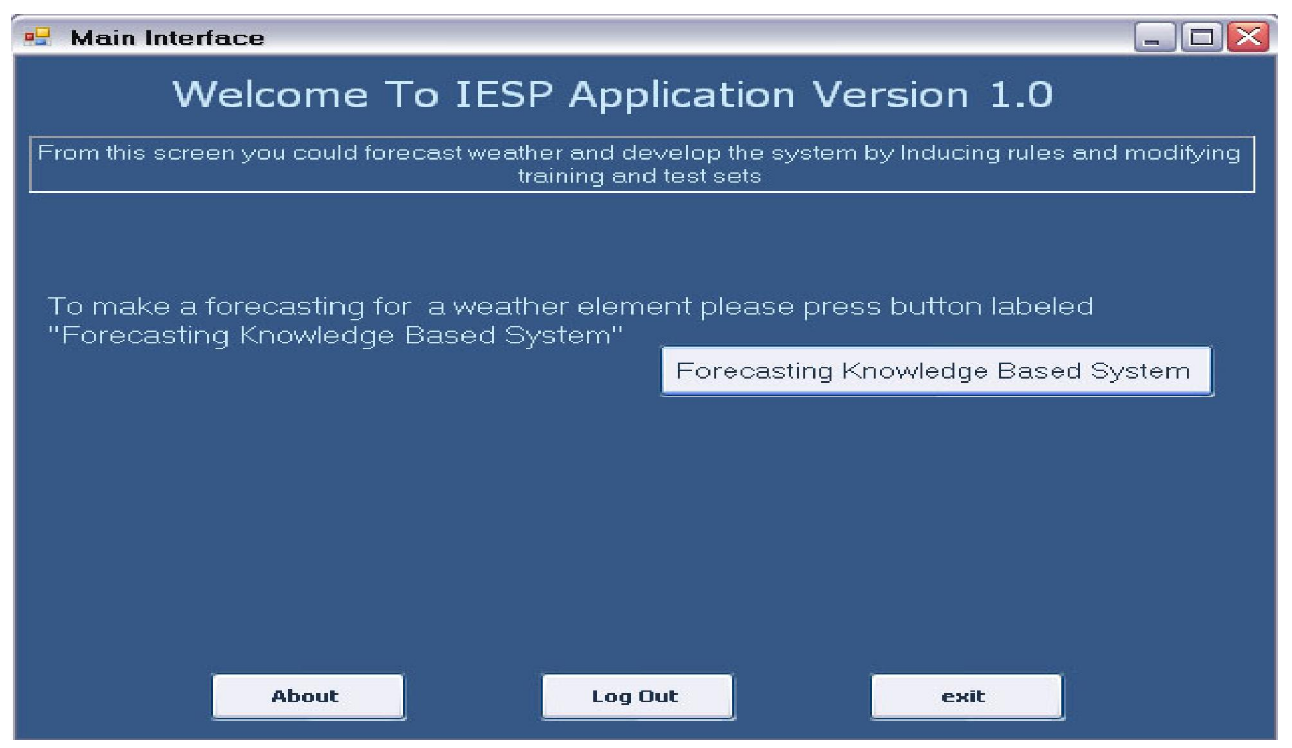
السطح البيني للمستخدم IESP

" السطح البيني المستخدم IESP " هو الآلية التي بواسطتها يتصل المستخدم مع النظام الخبيــر. ويجمــع " السطح البيني للمستخدم IESP " معطيات المدخلات بطريقة بسيطة وسريعة بالعمل مع مربعات الكومبــو. ويطلب السطح البيني للمستخدم من المستخدم أن يختار الحقيقة المناسبة من بين عدة بــدائل موجــودة فــي مربع الكومبو. ويتألف من الشاشات التالية:

شاشة المقدمة :

- **الشاشة الأولى "Main Interface"**

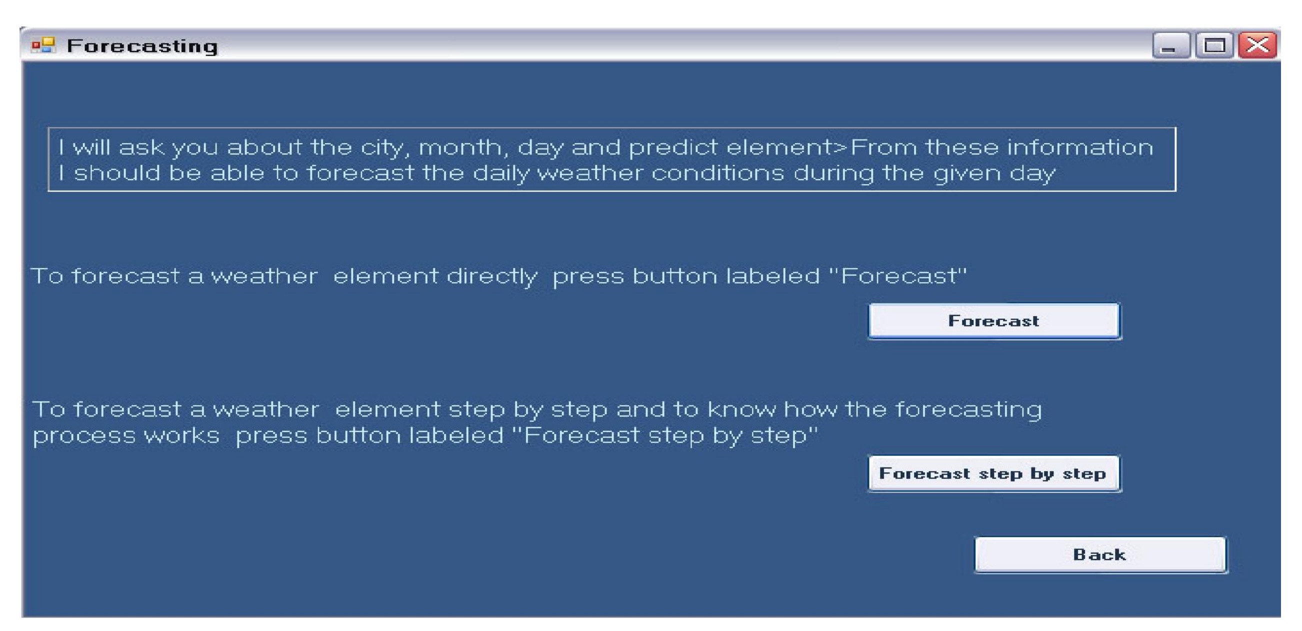
يمكن أن يدعم كل نظام خبير بشاشة مقدمة [١٢]. عند الحد الأدنى، تعطي هذه الشاشة نص ترحيبي للمستخدم. وأيضاً تخبر المستخدم عن الغرض العام للنظام. في IESP، تخبر شاشة المقدمة أن النظام يتنبأ بالطقس خلال كل الفصول وأيضاً تخبر المستخدم حول كيفية إنجاز النظام التنبؤ بالطقس. في IESP، لقد اخترنا أن نشرح أن النظام يستكشف اسم المدينة واسم الشهر واليوم والعنصر المطلوب للوصول إلى نتيجة منطقية. وفي IESP، توجد ثلاث دوال تقليدية للتحكم في تقدم المستخدم في شاشة المقدمة: الدالة الأولى "Forcasting Knowledge Based System" تكون مسؤولة عن بدء جلسة الحوار، والدالة الثانية "log Out" هي لإعادة بدء جلسة الحوار، والدالة "Exit" هي دالة الخروج من السطح البيني للمطور.



الشكل ٥-١١ شاشة المقدمة الأولى "Main Interface"

- **الشاشة الثانية "Forecasting"**

الشاشة "Forecasting" تحتوي ثلاث دوال، الدالة الأولى تمكن المستخدم من الحصول على نتيجة التنبؤ بشكل مباشر. والدالة الثانية تكون مسؤولة عن شرح خطوات التنبؤ للمستخدم. والثالثة للعودة إلى شاشـــة المقدمة الأولى.كما هو موضح بالشكل ٥-١٢.



الشكل ٥-١٢ شاشة المقدمة الثانية "Forecasting"

كما نرى تحتوي هذه الشاشة على زرين، الزر الأول "Forecast" يكون مســـؤول عـــن ظهـــور الشاشة "Direct Prediction" الموضحة في الشكل ٥-١٣، أما الـــزر الثـــاني "Forecast Step by Step" يكون مسؤول عن ظهور شاشة الأسئلة "Question Display".

الشكل ٥-١٣ الشاشة "Direct Prediction"

(ب) شاشة الأسئلة "Question Display"

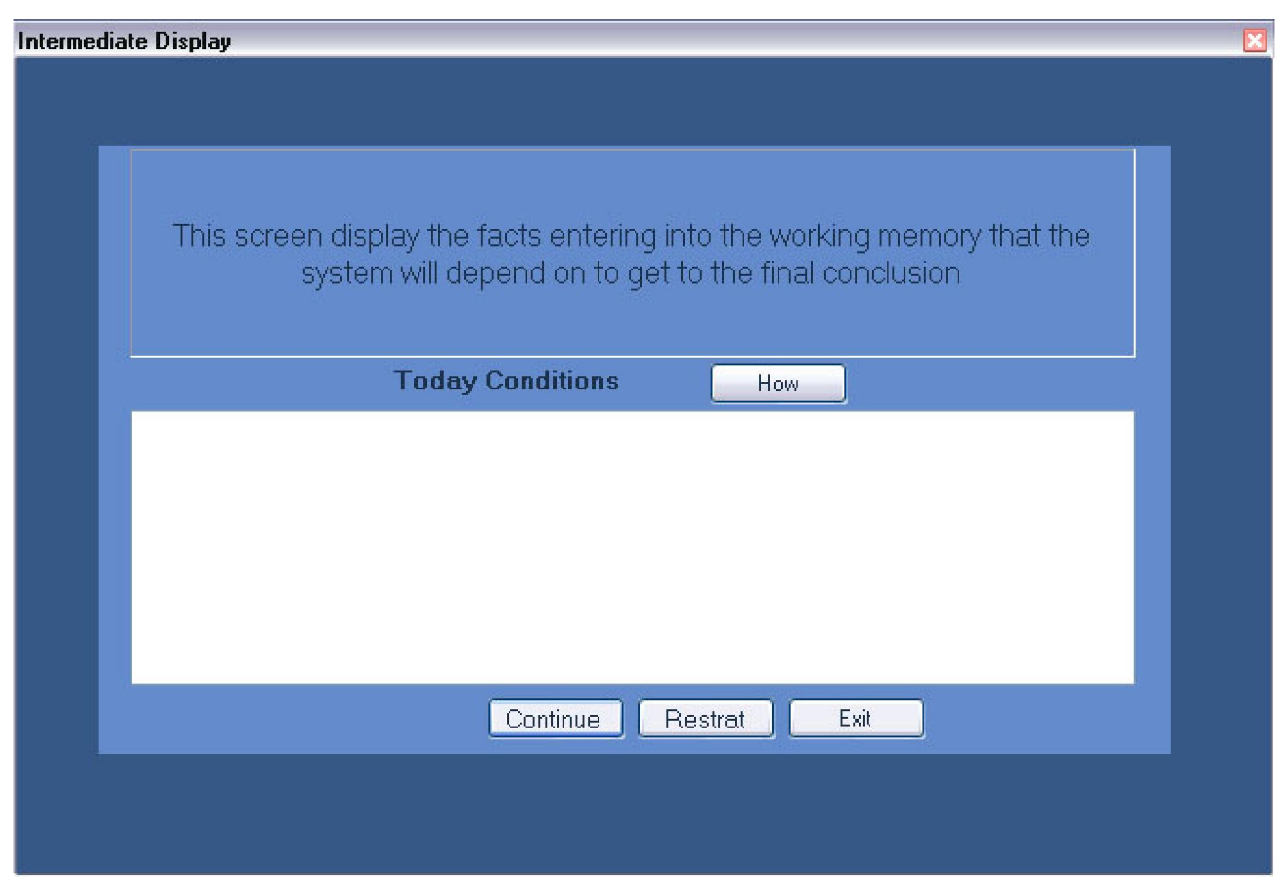
تستخدم شاشة الأسئلة للحصول على معلومات عن المشكلة من المستخدم. ويملك هذا النوع من الشاشات ثلاث أجزاء أساسية: جزء نصي يحتوي على السؤال، وجزء لإدخال الجواب، ومقطع للتحكم [١٢]. ويكون أحد الخيارات التي أخذناها بعين الاعتبار إضافة مقطع يقدم إشارة ما لماذا النظام يسأل السؤال. في IESP، تكون الدالة "Why?" مسؤولة عن التبرير، حيث تعرض هذه الدالة نص للمستخدم لكي يشرح لماذا يسأل النظام سؤال ما. في IESP، تحتوي " شاشة الأسئلة " على أربعة أسئلة. يسأل السؤال الأول عن المدينة، والثاني عن الشهر، والثالث عن اليوم، والرابع عن العنصر المطلوب. وتزود شاشة الأسئلة المستخدم بخيار "Restart" لاستمرار النظام، أو الخيار "Exit" يستخدم للخروج من النظام. وتوضح هذه الشاشة في الشكل ٥-١٤.

الشكل ٥-١٤ شاشة الأسئلة "Question Display"

(ج) الشاشة المرحلية " Intermediate Display "

تظهر هذه الشاشة بعد الضغط على الدالة "Predict"، يكون الغرض من الشاشة المرحلية إظهار للمستخدم المعالم الهامة التي وصل إليها النظام [١٢]. في IESP، تعرض هذه الشاشة الحقائق الجديدة (أي أنماط الطقس العامة) الداخلة إلى الذاكرة المستخدمة. ويكون أحد الخيارات الذي أخذناها بعين الاعتبار إضافة مقطع يقدم تبرير ما عن كيفية الوصول إلى النتيجة المرحلية. وفي IESP، تكون الدالة "How" مسؤولة عن التبرير، حيث تعرض هذه الدالة نص للمستخدم يشرح كيف وصل النظام إلى هذه الحقائق. وفي IESP، توجد ثلاثة دوال أساسية تقدم للمستخدم في "الشاشة المرحلية ": دالة للاستمرار، ودالة

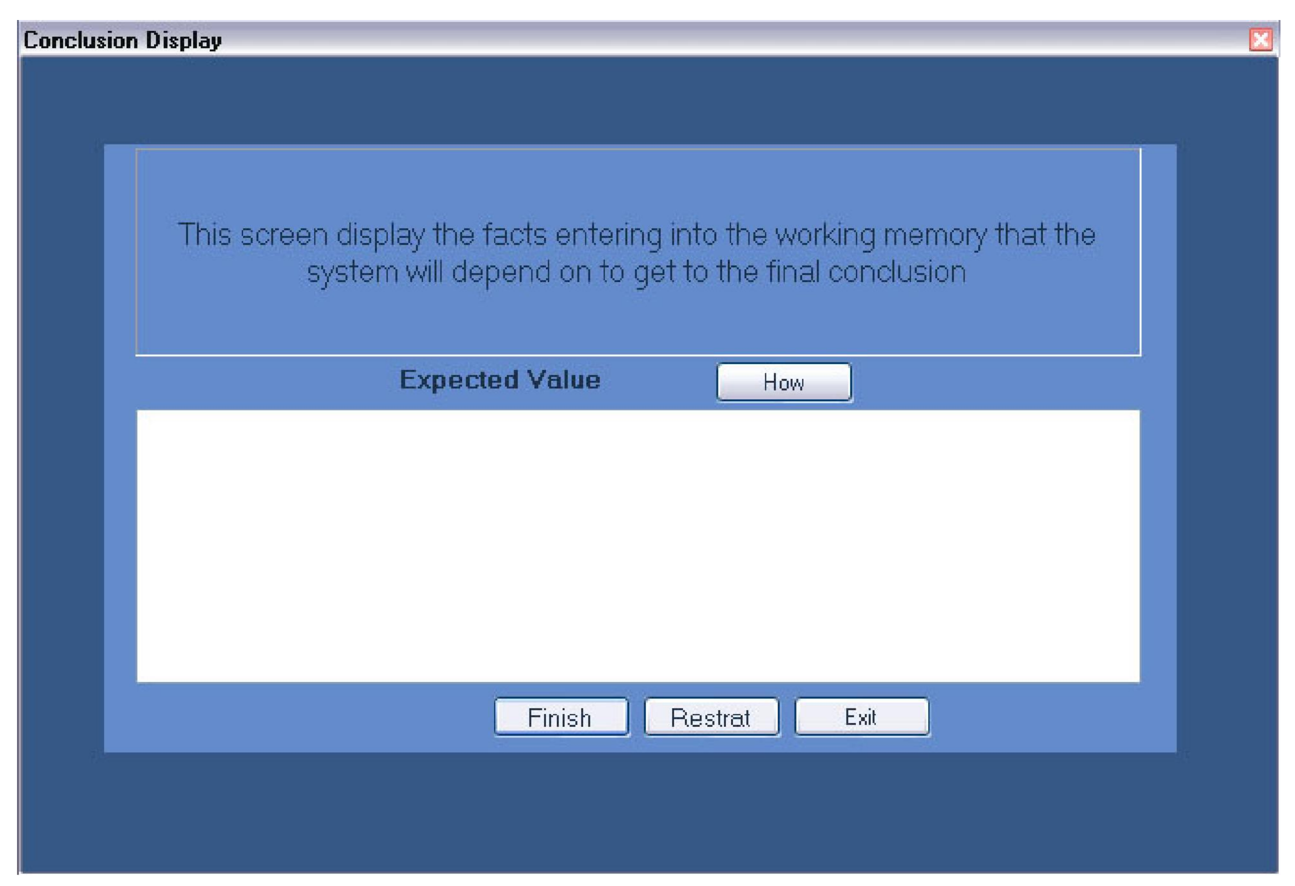
للخروج، ودالة لإعادة البدء. فتتابع دالة الاستمرار جلسة التحاور وتسمح دالة إعادة البدء للمستخدم لإعــادة بدء جلسة الحوار لتجريب أجوبة مختلفة. وتسمح دالة الخروج للمستخدم أن ينهي جلسة الحوار. وتوضـــح هذه الشاشة في لشكل ٥-١٥.



الشكل ٥-١٥ الشاشة المرحلية " Intermediate Display "

(د) شاشة النتيجة المنطقية " Conclusion Display "

تظهر هذه الشاشة بعد الضغط على الدالة "Continue"، تعرض شاشة النتيجة المنطقية للمســتخدم نتــائج النظام [١٢]. في IESP، تعرض هذه الشاشة نتائج التنبؤ بــالطقس. وتكــون العلامــة التجاريــة للــنظم الخبيرقدرتها على شرح كيف تم الوصول إلى النتيجة المنطقية النهائية. وفي IESP، تكون هذه الشروحات (التوضيحات) ضرورية للمستخدم لكي يقبل النتيجة المنطقية النهائية. وفي IESP، تكون الدالة " HOW? " مسؤولة عن الشروحات، حيث تعرض هذه الدالة نص للمستخدم يشرح كيــف وصــل النظــام للنتيجــة المنطقية النهائية. وتوجد دالة للتحكم في شاشة النتيجة المنطقية لإعادة البدء. وتوضح هذه الشاشة في الشكل ٥-١٦.



الشكل ٥-١٦ شاشة النتيجة المنطقية " Conclusion Display "

٥-٣-٢-٢ مجموعات التدريب والاختبار

مجموعة التدريب هي قاعدة معطيات تحتوي على معطيات التدريب التي تستخدم من قبل DBROUGH لاستقراء القواعد لنظامنا.

في IESP، تكون مجموعة التدريب من شكل ملفات ODBC. ويكون الشكل لملف معطيات التدريب مشابهة للملف المستخدم من قبل ملفات أكسس. ويملك كل سجل لمجموعة التدريب نفس الهيكل (نفس البنية) ويتألف من عدد من أزواج صفة – قيمة. وفي IESP، توجد مجموعة تدريب واحدة.

ومجموعة الاختبار هي قاعدة معطيات تحتوي على معطيات الاختبار التي تستخدم لاختبار نظامنا.

في IESP، تكون مجموعة الاختبار من شكل ملفات ODBC. ويكون الشكل لملف معطيات الاختبار مشابهاً للملف المستخدم من قبل ملفات الأكسس. ويملك كل سجل لمجموعة الاختبار نفس الهيكل ويتألف من عدد من الأزواج صفة– قيمة. في IESP، توجد ٦ مجموعات اختبار: مجموعة اختبار درجة الحرارة، ومجموعة اختبار السحب ومجموعة اختبار الرطوبة النسبية، ومجموعة اختبار سرعة الرياح، ومجموعة اختبار الضغط الجوي، ومجموعة اختبار ضغط بخار الماء.

أخيراً تكون كلاً من مجموعات التدريب والاختبار ديناميكية. هذا يعني، يمكننا إضافة وتعديل وحذف سجلات.

٥-٣-٢-٣ نظام اكتساب المعرفة الفرعي

نظام اكتساب المعرفة الفرعي هو طريقة آلية للمستخدم لإدخال المعرفة إلى النظام بدلاً من جعل مهندس المعرفة يشفر صراحة المعرفة [١٩].

في IESP، يكون نظام اكتساب المعرفة الفرعي نظام DBROUGH. تم عرض التفاصيل الكاملة لهذا النظام الفرعي سابقاً. باختصار، يستطيع هذا النظام الفرعي أن يؤدي مهمتين: (آ) اكتشاف أنماط الطقس العامة باستخدام خوارزمية استقراء تحليل الانحدار الموجه بالصفة، و(ب) واكتشاف القواعد باستخدام نظرية المجموعات التقريبية.

عموماً، تستطيع هذه الأداة أن تتعلم باستقراء القواعد من خلال الأمثلة ويمكن أن تولد القواعد آلياً. ومن ناحية ثانية، تكون الأمثلة عادة من أنواع معطيات جدولية كما شرح سابقاً.

٥-٣-٢-٤ قاعدة المعرفة

تحتوي قاعدة المعرفة لنظام IESP على خبرة المجال، أي القواعد التي حصل عليها من سجلات الطقس التاريخية بتطبيق نظام اكتساب المعرفة الفرعي المشروح أعلاه. وتستخدم قواعد"IF-THEN" البسيطة لتمثيل المعرفة في شكل قواعد كثيرة.

قسمت قاعدة المعرفة لنظام IESP إلى ٦ قواعد معرفة: قاعدة معرفة للتنبؤ بدرجة الحرارة، وقاعدة معرفة للتنبؤ بالسحب، وقاعدة معرفة للتنبؤ بالضغط الجوي، وقاعدة معرفة للتنبؤ بضغط بخار الماء ، وقاعدة معرفة للتنبؤ بالرطوبة، وقاعدة معرفة للتنبؤ بسرعة الرياح.

وتخزن كل قاعدة معرفة في ملف مستقل. وتكون قواعد المعرفة ديناميكية. هذا يعني، يمكننا إضافة وتعديل وحذف قواعد.

بعض قواعد قاعدة معرفة التنبؤ بدرجة الحرارة هي:

(a3=normal) ∧ (a4=moderate breeze) ∧ (a6=high) → (a1=slightly cold) (CF=0. 57)

(a3=normal) ∧ (a4=strong breeze) ∧ (a6=high) → (a1=warm) (CF=0. 67)

بعض قواعد قاعدة معرفة التنبؤ بالسحب هي:

(a3=moderate) ∧ (a4=moderate breeze) ∧ (a5=relatively high) ∧ (a6=high) → (a2=partly) (CF=0. 50)

(a3=relatively high) ∧ (a4=strong breeze) ∧ (a5=high)(a6=moderate) → (a2=fairly) (CF=1)

بعض قواعد قاعدة معرفة التنبؤ بالرطوبة النسبية هي:

(a1=slightly cold) ∧ (a2=partly) ∧ (a4=moderate breeze) ∧ (a5=relatively high)(a6=high) → (a3=moderate) (CF=0. 60)

(a1=slightly cold) ∧ (a2=fairly) ∧ a4=moderate breeze) ∧ (a5=relatively high) ∧ (a6=high) → (a3=normal) (CF=0. 67)

بعض قواعد قاعدة معرفة التنبؤ بسرعة الرياح هي:

(a1=slightly cold)∧(a2=partly)∧(a3=moderate)∧(a5=relatively high)∧(a6=high) → (a4=moderate breeze) (CF=1. 00)
(a1=slightly cold)∧(a2=fairly)∧(a3=normal)∧(a5=relatively high)∧(a6=high) → (a4=light breeze) (CF=0. 67)

بعض قواعد قاعدة معرفة التنبؤ بالضغط الجوي هي:

(a2=partly)∧(a3=moderate)∧(a4=moderate breeze)∧(a6=high) → (a5=relatively high) (CF=1. 00)
(a2=fairly)∧(a3=moderate)∧(a4=strong breeze)∧(a6=high) → (a5=high) (CF=0. 75)

بعض قواعد قاعدة معرفة التنبؤ بضغط بخار الماء هي:

(a1=slightly cold)∧(a2=partly)∧(a3=moderate)∧(a4=moderate breeze)∧(a5=relatively high) → (a6=high) (CF=0. 75)
(a1=slightly cold)∧(a2=fairly)(a3=relatively low)∧(a4=moderate breeze)∧ (a5=relatively high) → (a6=moderate) (CF=0. 50)

حيث يشير CF إلى عامل التأكد، الذي يمثل القياس العددي للثقة في صحة القاعدة. ويتراوح من القيمة صفر(عدم الثقة على الإطلاق) إلى القيمة واحد (الثقة الكاملة).

٥-٣-٢-٥ الذاكرة المستخدمة

تكون الذاكرة المستخدمة جزء من النظام الخبيروالتي تحتوي على حقائق المشكلة المكتشفة أثناء جلسة التحاور.

أثناء التحاور مع النظام IESP، يدخل المستخدم المعلومات عن المشكلة الحالية إلى الذاكرة المستخدمة. وعلاوة على ذلك، يستفيد النظام IESP من المعلومات المحتواة في التخزينات الخارجية، أي ملفات ODBC، حيث يحمل النظام بعض من هذه المعلومات وينفذ عليها خوارزمية " استقراء تحليل الانحدار الموجه بالصفة " التي تستنتج أنماط الطقس العامة. ومن ثم يدخل النظام هذه الحقائق الجديدة إلى الذاكرة المستخدمة. وبعد ذلك، يطابق النظام هذه المعلومات مع المعرفة المحتواة في قاعدة المعرفة للوصول إلى النتيجة المنطقية النهائية.

عموماً، تحتوي " الذاكرة المستخدمة " كل المعلومات حول المشكلة التي تزود إما بواسطة المستخدم أو المستنتجة بواسطة النظام.

٥-٣-٢-٦ آلية الاستدلال

تكون آلية الاستدلال، أيضاً تدعى محرك الاستدلال، المعالج في النظام الخبير الذي يقوم بالاستدلال حول المشكلة بمطابقة الحقائق المحتواة في الذاكرة المستخدمة مع معرفة المجال المحتواة في قاعدة المعرفة.

إنها تحدد استراتيجية التفكير (استنتاج المنطقي) المستخدمة وكيف يعالج عدم التأكد. في الواقع، تكون آلية الاستدلال الذكاء الذي يسمح للنظام الخبيرلأن يستنتج النتائج المنطقية بناءً على الخبرة أو المعرفة الملائمة لمجال المشكلة. وتستخدم آلية الاستدلال أسلوب التسلسل المتقدم مع خليط تتابعي ومتوازي من عوامل الثقة. ويبنى التسلسل المتقدم على فكرة فحص كل قاعدة باتجاه الأمام، باحثاً أولاً عن المقدمة المنطقية للقاعدة. ويتم تجاهل النتيجة المنطقية حتى تصبح المقدمة المنطقية صحيحة. وعندما تصبح المقدمة المنطقية للقاعدة صحيحة، تطلق القاعدة ومن ثم تأخذ الأفعال في نتيجتها المنطقية. ويوضح الشكل ٥-١٧ مخطط التدفق للتسلسل المتقدم لنظامنا. كما هو موضح بالشكل، يأخذ النظام كل قاعدة بالتتابع ويختبرها ليرى إذا مقدمتها المنطقية تعرض في الذاكرة المستخدمة. وعندما يجد النظام تطابقات لكل المقدمات المنطقية، يضع النتيجة المنطقية للقاعدة في الذاكرة المستخدمة. ومن ثم يدور ويكرر العملية. بهذا الشكل يكتسب النظام معلومات جديدة عن المشكلة التي يستخدمها في استنتاج منطقي آخر. وعندما لا توجد قواعد جديدة يمكن أن تطلق، يتوقف النظام.

تدير آلية الاستدلال النظام IESP عدم التأكد المرتبط مع النتائج المنطقية من خلال استخدام عام التأكد (CF) من نمط مايسن، العداد الذي يعكس مستوى الاعتقاد الصافي في فرضية بمعرفة المعلومات المتاحة [١٩، ١٧، ٩].

وتوضح الطريقة المستخدمة من قبل نظامنا في تحديد مستوى الثقة للنتيجة المنطقية النهائية بالصورة التالية:

بفرض أن لدينا القاعدة التالية:

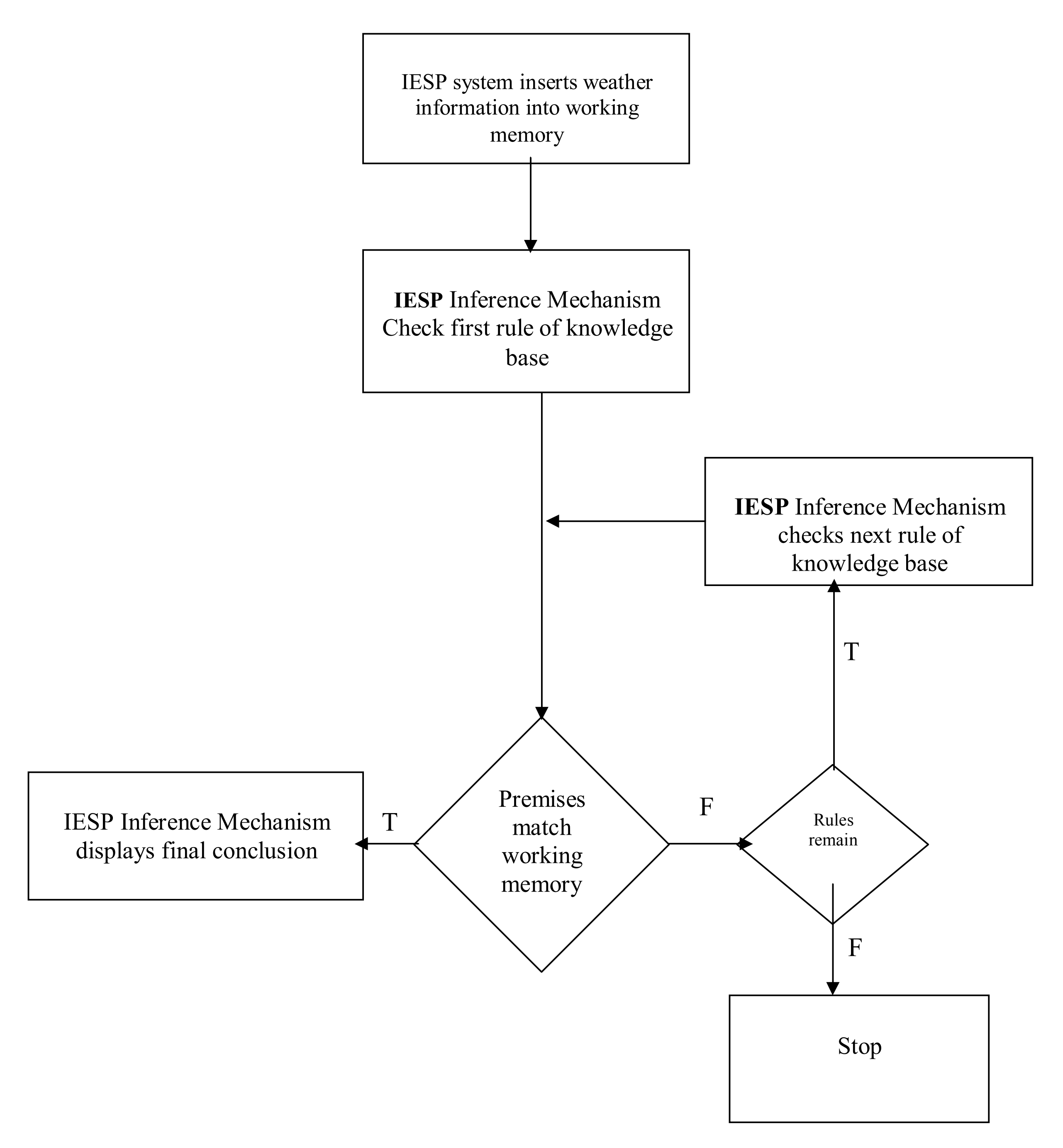
(a3=normal) $\wedge$ (a4=moderate breeze) $\wedge$ (a6=high) $\rightarrow$ (a1=slightly cold) (CF=0. 57)

افرض أن CF (Humidity)، و CF (Wind speed)، و CF (Vapor pressure)، و CF (Temperature) هي عوامل التأكد للرطوبة النسبية وسرعة الرياح وضغط بخار الماء ودرجة الحرارة، على التوالي. عندئذ

CF (Temperature)=min (CF (Humidity), CF (Wind speed), CF (Vapor pressure)) *CF,

حيث CF هو عامل التأكد للقاعدة المناظرة (المقابلة).

وتستخدم آلية الاستدلال لنظام IESP استراتيجية أولوية القاعدة لحل الخلاف (التضارب) حيث إنها تطلق القاعدة ذات عامل التأكد الأكبر.



الشكل ٥-١٧ مخطط التدفق المتسلسل

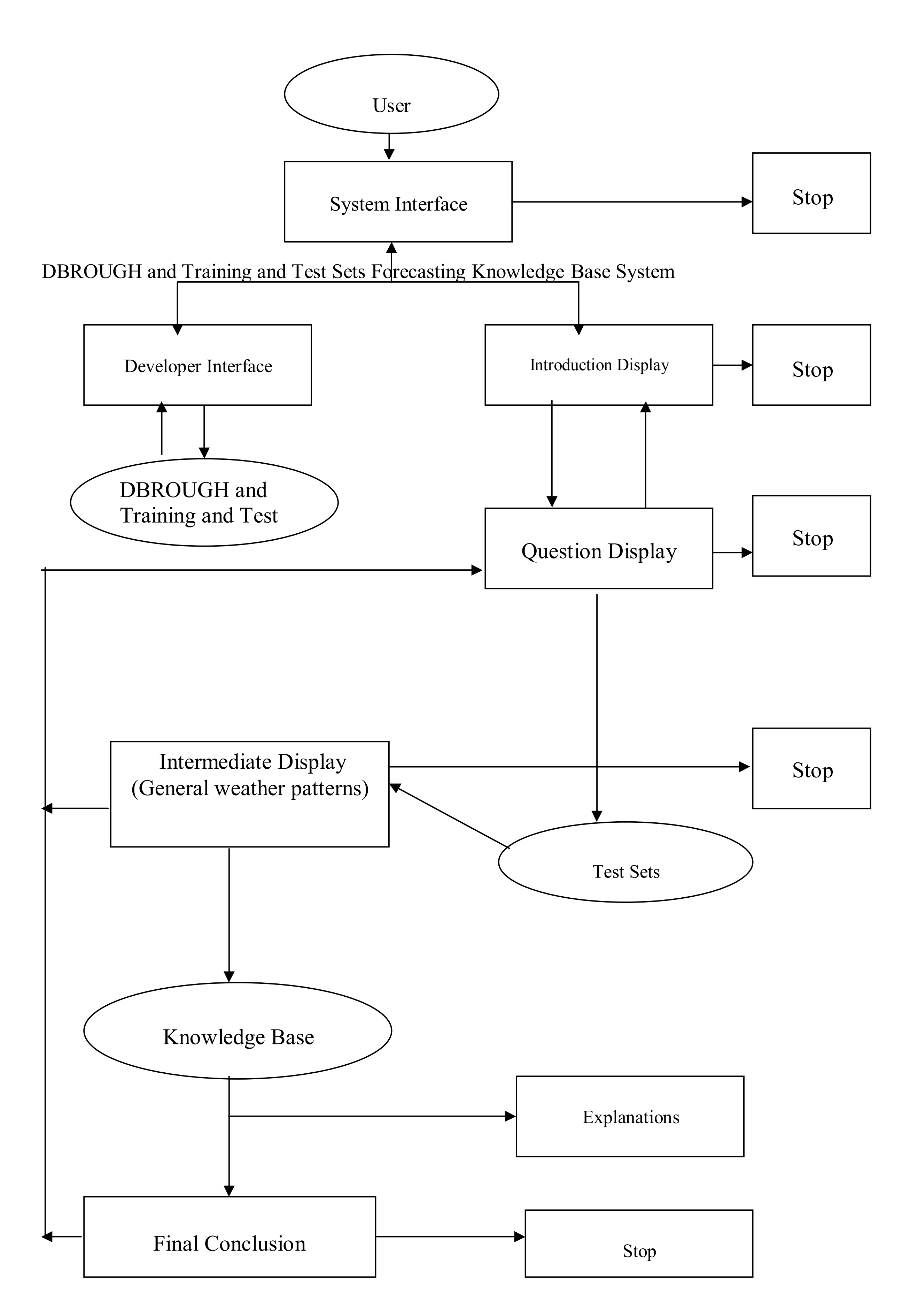
٥-٣-٢-٧ المبرر

يشرح المبرر، يدعى أيضاً نظام التوضيح الفرعي للمستخدم، الاستنتاج المنطقي للنظام [٩]. يحتوي نظام IESP نظام فرعي مصمم ليشرح للمستخدم لماذا النظام يسأل سؤال وكيف يصل النظام إلى نتائجه. ويقدم هذا النظام الفرعي فائدة لكل من مطور النظام ومستخدم النظام. فيمكن أن يستخدمه المطور في اكتشاف الأخطاء في معرفة النظام. ويستفيد المستخدم من الشفافية المقدمة في عملية الاستنتاج المنطقي للنظام. وعموماً، يكون السؤالان المدعومان من قبل نظام IESP استفساران "WHY " و"HOW ". ويكون

الغرض من استفسار "WHY" توضيحات وشروحاً لسبب احتياج النظام إلى أجراء معينة من المعلومـــات. في حين يكون الغرض من الاستفسار "HOW" تبرير نتائج النظام. وأثناء مرحلة التجاور، وبينما يطلـــب من المستخدم أن يقدم معلومات إلى النظام، يكون لديه الفرصة لأن يسأل لماذا، أي لماذا يسأل سؤال معـــين من قبل النظام الخبير؟ ويجيب النظام بعرض مربع رسالة يشرح لماذا يسأل النظام السؤال. وأيضاً، بعد أن يصل النظام إلى بعض النتائج، يستطيع أن يسأل المستخدم كيف، أي، كيف وصل إلى نتيجة منطقية معينة؟ ويجاب على الاستفسارات "HOW" بإعطاء تتبع لعملية الاستنتاج المنطقـــي التـــي قـــادت إلـــى النتيجـــة المنطقية. بكلمات أخرى، يعرض نظام IESP في مربع تحرير القاعدة أو مجموعة القواعـــد التـــي تحـــدد النتيجة التي وصل إليها.

٥-٤ تشغيل النظام

يوضح مخطط التدفق الرئيسي لنظام IESP في الشكل ٥-١٨ إذا أراد المستخدم الذي دخـــل إلـــى النظـــام الدخول إلى DBROUGH ومجموعات التدريب والاختبار، عندئذ ينقله النظام إلى السطح البيني للمطـــور الذي منه يدخل إلى DBROUGH ومجموعات التدريب والاختبار. وإذا أراد المستخدم الذي دخـــل إلـــى النظام الدخول إلى النظام الخبير، عندئذ ينقله النظام إلى شاشة المقدمة التي تظهر عبارات ترحيبية وشـــرح مختصر عن كيفية عمل النظام، وكيف يستطيع المستخدم أن يتعامل معه. ومن ثم ينقله النظام إلـــى شاشـــة الأسئلة. وبعد أن يجيب المستخدم على الأسئلة، يعطي النظام التحكم إلى DBROUGH (أي نظام اكتساب المعرفة الفرعي)، الذي يكتشف أنماط الطقس العامة بناءً على المعطيات المناخية المرصودة في مجموعات الاختبار. وبعد ذلك يعطي DBROUGH التحكم إلى النظام الذي يعرض هذه الحقائق الجديدة في الشاشـــة المرحلية. ومن ثم يعطي النظام التحكم إلى آلية الاستدلال التي تأخـــذ النتـــائج المعروضـــة فـــي الشاشـــة المرحلية، وتبحث في قاعدة المعرفة، وتعرض النتيجة المنطقية النهائية في شاشة النتيجة المنطقية. ومن ثـــم تعطي آلية الاستدلال التحكم للنظام لبدء التحاور مرة أخرى أو للخروج من النظام.



الشكل ٥– ١٨ مخطط التدفق الرئيسي لنظام IESP

يعرض التحاور بين النظام الخبير والمستخدم والذي يوضح تشغيل النظام في الملحق A.

مراجعة التحاور

من هذا التحاور، نلاحظ أن نظامنا كان لطيفاً طوال جلسة التحاور. فتعلم شاشة المقدمة المستخدم ماذا سينجز النظام وتعطيه تبصر عن كيفية عمله. وتطرح الأسئلة في شكل سهل التتبع في شاشة الأسئلة مع تبرير معين عندما ينقر المستخدم على الزر "WHY?". وتعطي الشاشة المرحلية أنماط الطقس العامة وتبرير ما عندما ينقر المستخدم عل الزر "HOW?". وتعطي شاشة النتيجة المنطقية التنبؤ النهائي وتبرير ما عندما ينقر المستخدم على الزر "HOW?".

٥-٥ ملاحظات ختامية

في هذا الفصل، بنينا نظام خبير استقرائي فريد للتنبؤ بالطقس، IESP.وهذا النظام نظام مبني على الاستقراء يستغل نموذج تحليل الانحدار ونظرية المجموعات التقريبية الموجه بالصفة لاكتشاف أنماط الطقس العامة وقواعد المعرفة من مجموعات التدريب والاختبار. ويملك نظامنا السمات التالية:

١. يحوي المعرفة في شكل قواعد التي تساعد المتنبئ بالتنبؤ بالطقس.
٢. يملك ٦ قواعد معرفة.
٣. يملك مجموعتي تدريب و٦ مجموعات اختبار.
٤. يفصل المعرفة عن التحكم.
٥. يملك آلية استدلال (محرك الاستدلال) تستخدم التسلسل المتقدم على طول خليط تتابعي ومتوازي من عوامل الثقة.
٦. يملك نظام اكتساب معرفة فرعي (DBROUGH) لاكتشاف أنماط الطقس العامة والقواعد من قواعد المعطيات الطقسية العلائقية (أي، مجموعات التدريب والاختبار)، حيث يوظف هذا النظام الفرعي نموذج تحليل الانحدار ونظرية المجموعات التقريبية لذلك.
٧. يسمح لمعرفتة لأن تعدل بسهولة بتنقيح أو بحذف أو بإضافة القواعد.
٨. يملك سطح بيني للنظام قوي ومرن، متضمناً السطح البيني للمستخدم والسطح البيني للمطور.
٩. يملك نظام توضيح فرعي (مبرر) يقدم شرح وتبرير لأفعال النظام (الاستفسارات HOW وWHY).
١٠. يدير عدم التأكد المرتبط بالنتائج المنطقية من خلال استخدام عوامل التأكد من نمط ما يسن والمنطق الضبابي.
١١. يكون قابل للنقل، هذا يعني إمكانية استخدام أكثر من نوع واحد من الحاسبات لتنفيذ النظام.

الفصل السادس

اختبار النظام

٦-١ مقدمة

بعد أن بنينا النظام، يكون هدفنا اختباره لنرى ما إذا كان يلبي المتطلبات والأهداف الأصلية التي لأجلها تم البناء. وتكون عملية الاختبار مرحلة هامة في تطوير النظام الخبير حيث تمكن عملية التغذية العكسية من الحدوث والتي بواسطتها تخدم التعليقات كأساس لتنقيات متكررة للنظام. وفعلياً، ووفقاً جاسشينج وآخرون، تختبر النظم الخبيرة أساساً للتأكد من دقة وفائدة النظام [١٧].

يقدم هذا الفصل " اختبار النظام " في المقطع ٦- ٢، ويختتم الفصل بملاحظات ختامية في المقطع ٦-٣.

٦-٢ اختبار النظام

تقارن مهمة الاختبار النتائج التي يتنبأ بها النظام مع دراسات الحالة الفعلية (الحقيقية) [١٧]. بكلمات أخرى، تهتم هذه المهمة باختبار النظام ببعض الحالات الحقيقية. وتكون مهمة الاختبار إما بواسطة خبير النطاق أو من خلال أدلة التنبؤ بالطقس كمصدر للمعرفة. ولكن. نظراً لاستخدامنا السجلات التاريخية كمصدر للمعرفة، يمكننا أن نعمل مباشرةً مع السجلات التاريخية من أجل حالات الاختبار.

٦-٢-١ اختبار النظام في تطبيق مناخي

في هذا المقطع الفرعي، سنختبر النظام في تطبيق مناخي. في هذا التطبيق، سنختبر النظام للتنبؤ الشهري بدرجات الحرارة، وكميات السحب، والرطوبة النسبية، وسرع الرياح، والضغوط الجوية، وضغوط بخار الماء. وكل ما يجب أن نفعله يكون الرجوع إلى السجلات التاريخية ونقارن هذه السجلات مع النتائج التي ينبأ بها النظام.

بعد الرجوع إلى السجلات التاريخية، وجدنا درجات الحرارة، وكميات السحب، والرطوبة النسبية، وسرع الرياح، والضغوط الجوية، وضغوط بخار الماء المرصودة في فصل الشتاء من ١٩٨٥ إلى ٢٠٠٧ كما هو موضح في الجداول ٦-١، ٦-٢، ٦-٣، ٦-٤، ٦-٥، ٦-٦، على التوالي. وبعد تنفيذ النظام في هذه السنوات، تنبأ النظام بالقيم الموضحة في الجداول المذكورة أعلاه.

تحليل النتائج

كما هو موضح في الجدول ٦-١، نلاحظ وجود تشابهات كاملة بين درجات الحرارة المرصودة والمتنبأ بها في بعض السنوات.

year	Month	Observed temperature	Predicted temperature
1997	1	Slightly cold	Slightly cold
1996	1	Slightly cold	Slightly cold
1995	1	Slightly cold	Slightly cold
1988	1	Slightly cold	Slightly cold
2003	1	Slightly cold	Slightly cold
2002	1	Slightly cold	Slightly cold
2001	1	Slightly cold	Slightly cold
1998	1	Slightly cold	Slightly cold
1995	2	Warm	Warm
1994	2	Slightly cold	Slightly cold
1991	2	Warm	Warm
1989	2	Slightly cold	Slightly cold
1986	2	Slightly cold	Slightly cold
2003	2	Warm	Warm
1998	2	Warm	Warm
2007	2	Warm	Warm
1997	12	Warm	Warm
1996	12	Warm	Warm
1995	12	Warm	Warm
1994	12	Moderate	Moderate
1992	12	Warm	Warm
1991	12	Warm	Warm
1990	12	Warm	Warm
1988	12	Warm	Warm
1986	12	Warm	Warm

الجدول ٦-١ وجود تشابهات كاملة بين درجات الحرارة المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-٢، نلاحظ وجود شبه تشابهات بين درجات الحرارة المرصودة والمتنبأ بها في بعض السنوات.

year	Month	Observed temperature	Predicted temperature
1994	1	Slightly cold	Slightly cold/Warm
1993	1	Slightly cold	Slightly cold/Warm
1992	1	Slightly cold	Slightly cold/Warm
1991	1	Slightly cold	Slightly cold/Warm
1990	1	Slightly cold	Slightly cold/Warm
1989	1	Warm	Slightly cold/Warm
1987	1	Slightly cold	Slightly cold/Warm
1986	1	Slightly cold	Slightly cold/Warm
1985	1	Slightly cold	Slightly cold/Warm
2004	1	Slightly cold	Slightly cold/Warm
2000	1	Slightly cold	Slightly cold/Warm
1999	1	Slightly cold	Slightly cold/Warm
2007	1	Slightly cold	Slightly cold/Warm
2006	1	Slightly cold	Slightly cold/Warm
2005	1	Warm	Slightly cold/Warm
1997	2	Slightly cold	Slightly cold/Warm
1996	2	Slightly cold	Slightly cold/Warm
1993	2	Slightly cold	Slightly cold/Warm
1988	2	Warm	Slightly cold/Warm
1987	2	Slightly cold	Slightly cold/Warm
1985	2	Slightly cold	Slightly cold/Warm
2004	2	Slightly cold	Slightly cold/Warm
2002	2	Warm	Slightly cold/Warm
2001	2	Warm	Slightly cold/Warm
2000	2	Slightly cold	Slightly cold/Warm
1999	2	Slightly cold	Slightly cold/Warm
2006	2	Slightly cold	Slightly cold/Warm
2005	2	Warm	Slightly cold/Warm
1993	12	Warm	Warm/Moderate
1989	12	Moderate	Warm/Moderate
1985	12	Warm	Warm/Moderate
2004	12	Moderate	Warm/Moderate
2003	12	Moderate	Warm/Moderate
2002	12	Warm	Warm/Moderate
2001	12	Warm	Warm/Moderate
1999	12	Warm	Warm/Moderate
1998	12	Warm	Warm/Moderate
2007	12	Warm	Warm/Moderate
2006	12	Warm	Warm/Moderate
2005	12	Warm	Warm/Moderate

الجدول ٦-٢ وجود شبه تشابهات بين درجات الحرارة المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-٣، نلاحظ وجود اختلافات واضحة بين درجات الحرارة المرصودة والمتنبأ بها في بعض السنوات.

year	Month	Observed temperature	Predicted temperature
1992	2	Slightly cold	Warm
1990	2	Warm	Slightly cold
1987	12	Moderate	Warm
2000	12	Warm	Moderate

الجدول ٦-٣ وجود اختلافات واضحة بين درجات الحرارة المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-٤، نلاحظ وجود تشابهات كاملة بين كميات السحب المرصودة والمتنبأ بهـا في بعض السنوات.

year	Month	Observed Cloud	Predicted Cloud
1997	1	Fairly	Fairly
1996	1	Fairly	Fairly
1995	1	Fairly	Fairly
1988	1	Partly	Partly
2003	1	Fairly	Fairly
2002	1	Fairly	Fairly
2001	1	Fairly	Fairly
1997	2	Fairly	Fairly
1995	2	Fairly	Fairly
1994	2	Fairly	Fairly
1986	2	Fairly	Fairly
1998	2	Fairly	Fairly
1997	12	Fairly	Fairly
1996	12	Fairly	Fairly
1995	12	Fairly	Fairly
1994	12	Partly	Partly
1991	12	Fairly	Fairly
1990	12	Fairly	Fairly
1989	12	Fairly	Fairly
1988	12	Fairly	Fairly
1987	12	Fairly	Fairly
1986	12	Fairly	Fairly
2004	12	Fairly	Fairly
2003	12	Partly	Fairly
1999	12	Fairly	Fairly
2007	12	Fairly	Fairly

الجدول ٦-٤ وجود تشابهات كاملة بين كميات السحب المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-٥، نلاحظ وجود شبه تشابهات بين كميات السحب المرصودة والمتنبأ بها في بعض السنوات.

year	Month	Observed Cloud	Predicted Cloud
1994	1	Fairly	Fairly/Partly
1993	1	Fairly	Fairly/Partly
1992	1	Fairly	Fairly/Partly
1990	1	Partly	Fairly/Partly
1989	1	Partly	Fairly/Partly
1987	1	Fairly	Fairly/Partly
1986	1	Fairly	Fairly/Partly
1985	1	Fairly	Fairly/Partly
2004	1	Fairly	Fairly/Partly
2000	1	Partly	Fairly/Partly
1999	1	Partly	Fairly/Partly
2007	1	Partly	Fairly/Partly
2006	1	Partly	Fairly/Partly
2005	1	Fairly	Fairly/Partly
1996	2	Fairly	Fairly/Partly
1993	2	Partly	Fairly/Partly
1991	2	Partly	Fairly/Partly
1988	2	Fairly	Fairly/Partly
1987	2	Fairly	Fairly/Partly
1985	2	Fairly	Fairly/Partly
2004	2	Partly	Fairly/Partly
2003	2	Partly	Fairly/Partly
2002	2	Partly	Fairly/Partly
2001	2	Partly	Fairly/Partly
2000	2	Partly	Fairly/Partly
1999	2	Fairly	Fairly/Partly
2007	2	Fairly	Fairly/Partly
2006	2	Fairly	Fairly/Partly
2005	2	Fairly	Fairly/Partly
1993	12	Partly	Fairly/Partly
1992	12	Fairly	Fairly/Partly
1985	12	Fairly	Fairly/Partly
2002	12	Fairly	Fairly/Partly
2001	12	Fairly	Fairly/Partly
1998	12	Fairly	Fairly/Partly
2006	12	Fairly	Fairly/Partly

الجدول ٦-٥ وجود شبه تشابهات بين كميات السحب المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-٦، نلاحظ وجود فروقات واضحة بين كميات الســحب المرصــودة والمتنبأ بها في بعض السنوات.

year	Month	Observed Cloud	Predicted Cloud
1991	1	Fairly	Partly
1998	1	Partly	Fairly
1990	2	Fairly	Partly
1989	2	Fairly	Partly
1992	12	Fairly	Partly
2003	12	Partly	Fairly
2000	12	Fairly	Partly
2005	12	Partly	Fairly

الجدول ٦-٦ وجود فروقات واضحة بين كميات السحب المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-٧، نلاحظ وجود تشابهات كاملة بين الرطوبة النســبية المرصــودة والمتنبأ بها في بعض السنوات.

year	Month	Observed Humidity	Predicted Humidity
1996	1	Moderate	Moderate
1988	1	Moderate	Moderate
2004	1	Normal	Normal
2007	1	Normal	Normal
1997	2	Normal	Normal
1995	2	Normal	Normal
1994	2	Moderate	Moderate
1992	2	Normal	Normal
1990	2	Normal	Normal
1989	2	Normal	Normal
1986	2	Normal	Normal
1995	12	Moderate	Moderate
1994	12	Moderate	Moderate
1990	12	Moderate	Moderate
1986	12	Normal	Normal
2001	12	Normal	Normal
2000	12	Moderate	Moderate

الجدول ٦-٧ وجود تشابهات كاملة بين الرطوبة النسبية المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-٨، نلاحظ وجود شبه تشابهات بين الرطوبة النسبية المرصودة والمتنبأ بها في بعض السنوات.

year	Month	Observed Humidity	Predicted Humidity
1997	1	Moderate	Moderate/Relatively High
1990	1	Normal	Relatively Low/Normal
1989	1	Normal	Relatively Low/Normal
1986	1	Moderate	Normal/Moderate
2003	1	Normal	Relatively Low/Normal
2002	1	Normal	Relatively Low/Normal
2000	1	Normal	Normal/Moderate
2006	1	Moderate	Normal/Moderate
2005	1	Normal	Normal/Moderate
1996	2	Moderate	Normal/Moderate
1991	2	Normal	Relatively Low/ Normal
1988	2	Normal	Normal/Moderate
1987	2	Normal	Relatively Low/ Normal
1985	2	Normal	Relatively Low/ Normal
2003	2	Normal	Relatively Low/ Normal
2001	2	Normal	Normal/Moderate
2000	2	Normal	Relatively Low/ Normal
1999	2	Normal	Relatively Low/ Normal
1998	2	Moderate	Normal/Moderate
2007	2	Normal	Relatively Low/ Normal
2006	2	Normal	Normal/Moderate
1997	12	Moderate	Normal/Moderate
1996	12	Moderate	Normal/Moderate
1991	12	Moderate	Normal/Moderate
1988	12	Moderate	Normal/Moderate
1985	12	Normal	Normal/Relatively High
2003	12	Normal	Normal/Relatively High
2005	12	Normal	Normal/Relatively High

الجدول ٦-٨ وجود شبه تشابهات بين الرطوبة النسبية المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-٩، نلاحظ وجود فروقات واضحة بين الرطوبة النسبية المرصـــودة والمتنبأ بها في بعض السنوات.

year	Month	Observed Humidity	Predicted Humidity
1995	1	Moderate	Relatively Low/Normal
1994	1	Moderate	Relatively Low/Normal
1993	1	Relatively High	Relatively Low/Normal
1992	1	Moderate	Relatively Low/Normal
1991	1	Normal	Moderate
1987	1	Moderate	Relatively Low/Normal
1985	1	Relatively Low	Normal
2001	1	Moderate	Relatively Low/Normal
1999	1	Moderate	Relatively Low/Normal
1998	1	Moderate	Relatively Low/Normal
1993	2	Moderate	Relatively Low/ Normal
2004	2	Moderate	Relatively Low/ Normal
2002	2	Relatively Low	Normal/Moderate
2005	2	Relatively Low	Normal/Moderate
1993	12	Moderate	Normal
1992	12	Normal	Moderate
1989	12	Moderate	Normal/Relatively High
1987	12	Normal	Moderate
2004	12	Moderate	Normal/Relatively High
2002	12	Moderate	Normal
1999	12	Moderate	Normal/Relatively High
1998	12	Moderate	Normal/Relatively High
2007	12	Moderate	Normal/Relatively High
2006	12	Relatively High	Normal/Moderate

الجدول ٦-٩ وجود فروقات واضحة بين الرطوبة النسبية المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-١٠، نلاحظ وجود تشابهات كاملة بين ســـرع الريـــاح المرصـــودة والمتنبأ بها في بعض السنوات.

year	Month	Observed Wind Speed	Predicted Wind Speed
1997	1	Strong Breeze	Strong Breeze
1991	1	Strong Breeze	Strong Breeze
1985	1	Moderate Breeze	Moderate Breeze

2004	1	Moderate Breeze	Moderate Breeze
2002	1	Moderate Breeze	Moderate Breeze
2001	1	Moderate Breeze	Moderate Breeze
1998	1	Moderate Breeze	Moderate Breeze
2007	1	Moderate Breeze	Moderate Breeze
1997	2	Strong Breeze	Strong Breeze
1995	2	Strong Breeze	Strong Breeze
1994	2	Strong Breeze	Strong Breeze
1992	2	Strong Breeze	Strong Breeze
1989	2	Strong Breeze	Strong Breeze
1986	2	Strong Breeze	Strong Breeze
1998	2	Moderate Breeze	Moderate Breeze
2007	2	Moderate Breeze	Moderate Breeze
1996	12	Tornado	Tornado
1992	12	Tornado	Tornado
2007	12	Moderate Breeze	Moderate Breeze

الجدول ٦-١٠ وجود تشابهات كاملة بين سرع الرياح المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-١١، نلاحظ وجود شبه تشابهات بين سرع الرياح المرصودة والمتنبأ بها في بعض السنوات.

year	Month	Observed Wind Speed	Predicted Wind Speed
1996	1	Strong Breeze	Moderate Breeze/ Strong Breeze
1994	1	Strong Breeze	Moderate Breeze/ Strong Breeze
1986	1	Strong Breeze	Moderate Breeze/ Strong Breeze
2000	1	Moderate Breeze	Moderate Breeze/Light Breeze
1999	1	Moderate Breeze	Moderate Breeze/Light Breeze
2006	1	Moderate Breeze	Moderate Breeze/Light Breeze
2005	1	Moderate Breeze	Moderate Breeze/Light Breeze
1996	2	Strong Breeze	Moderate Breeze/ Strong Breeze
1988	2	Strong Breeze	Moderate Breeze/ Strong Breeze
1987	2	Strong Breeze	Strong Breeze/Tornado
2002	2	Moderate Breeze	Moderate Breeze/ Strong Breeze
2001	2	Moderate Breeze	Moderate Breeze/ Strong Breeze
2006	2	Moderate Breeze	Moderate Breeze/ Strong Breeze
2005	2	Moderate Breeze	Moderate Breeze/ Strong Breeze
1994	12	Hurricane	Moderate Breeze/Hurricane
1990	12	Tornado	Moderate Breeze/Tornado
1986	12	Tornado	Moderate Breeze/Tornado
1985	12	Moderate Breeze	Light Breeze/Moderate Breeze
2004	12	Moderate Breeze	Moderate Breeze/Tornado
2003	12	Moderate Breeze	Moderate Breeze/Tornado
2000	12	Light Breeze	Light Breeze/Hurricane

1999	12	Moderate Breeze	Moderate Breeze/Tornado
1998	12	Moderate Breeze	Light Breeze/Moderate Breeze
2006	12	Moderate Breeze	Light Breeze/Moderate Breeze
2005	12	Moderate Breeze	Light Breeze/Moderate Breeze

الجدول ٦-١١ وجود شبه تشابهات بين سرع الرياح المرصودة والمتنبأ بها في بعض السنوات.

كما هو موضح في الجدول ٦-١٢، نلاحظ وجود فروقات واضحة بين سرع الريــاح المرصــودة والمتنبأ بها في بعض السنوات.

year	Month	Observed Wind Speed	Predicted Wind Speed
1995	1	Strong Breeze	Moderate Breeze
1993	1	Strong Breeze	Moderate Breeze/Light Breeze
1992	1	Strong Breeze	Moderate Breeze/Light Breeze
1990	1	Strong Breeze	Moderate Breeze/Light Breeze
1989	1	Strong Breeze	Moderate Breeze/Light Breeze
1988	1	Strong Breeze	Moderate Breeze
1987	1	Strong Breeze	Moderate Breeze/Light Breeze
2003	1	Light Breeze	Moderate Breeze
1993	2	Strong Breeze	Light Breeze/Tornado
1991	2	Tornado	Moderate Breeze
1990	2	Tornado	Moderate Breeze
1985	2	Moderate Breeze	Strong Breeze/Tornado
2004	2	Moderate Breeze	Strong Breeze/Tornado
2003	2	Light Breeze	Moderate Breeze
2000	2	Moderate Breeze	Strong Breeze/Tornado
1999	2	Moderate Breeze	Strong Breeze/Tornado
1997	12	Tornado	Moderate Breeze/Hurricane
1995	12	Tornado	Light Breeze/Moderate Breeze
1993	12	Tornado	Moderate Breeze
1991	12	Tornado	Light Breeze/Moderate Breeze
1989	12	Hurricane	Moderate Breeze/Tornado
1988	12	Tornado	Light Breeze/Moderate Breeze
1987	12	Hurricane	Moderate Breeze/Tornado
1986	12	Tornado	Moderate Breeze/Tornado
2002	12	Light Breeze	Moderate Breeze
2001	12	Light Breeze	Moderate Breeze

الجدول ٦-١٢ وجود فروقات واضحة بين سرع الرياح المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-١٣، نلاحظ وجود تشابهات كاملة بين الضغط الجــوي المرصــودة والمتنبأ بها في بعض السنوات.

year	Month	Observed Pressure	Predicted Pressure
1997	1	High	High
1995	1	Relatively High	Relatively High
1992	1	Relatively High	Relatively High
1990	1	Relatively High	Relatively High
1989	1	Relatively High	Relatively High
1986	1	Relatively High	Relatively High
1985	1	Relatively High	Relatively High
2004	1	Relatively High	Relatively High
2000	1	Relatively High	Relatively High
1999	1	Relatively High	Relatively High
2007	1	Relatively High	Relatively High
2006	1	Relatively High	Relatively High
2005	1	Relatively High	Relatively High
1997	2	Relatively High	Relatively High
1996	2	Relatively High	Relatively High
1995	2	Relatively High	Relatively High
1994	2	Relatively High	Relatively High
1992	2	Relatively High	Relatively High
1991	2	Relatively High	Relatively High
1990	2	Relatively High	Relatively High
1989	2	Relatively High	Relatively High
1988	2	Relatively High	Relatively High
1987	2	Relatively High	Relatively High
1986	2	Relatively High	Relatively High
1985	2	Relatively High	Relatively High
2004	2	Relatively High	Relatively High
2003	2	Relatively High	Relatively High
2002	2	Relatively High	Relatively High
2001	2	Relatively High	Relatively High
2000	2	Relatively High	Relatively High
1999	2	Relatively High	Relatively High
1998	2	Relatively High	Relatively High
2007	2	Relatively High	Relatively High
2006	2	Relatively High	Relatively High

2005	2	Relatively High	Relatively High
1997	12	Relatively High	Relatively High
1996	12	Relatively High	Relatively High
1995	12	Relatively High	Relatively High
1994	12	Relatively High	Relatively High
1993	12	Relatively High	Relatively High
1991	12	Relatively High	Relatively High
1990	12	Relatively High	Relatively High
1989	12	Relatively High	Relatively High
1988	12	Relatively High	Relatively High
1987	12	Relatively High	Relatively High
1986	12	Relatively High	Relatively High
1985	12	Relatively High	Relatively High
2004	12	Relatively High	Relatively High
2002	12	Relatively High	Relatively High
2001	12	Relatively High	Relatively High
2000	12	Relatively High	Relatively High
1999	12	Relatively High	Relatively High
1998	12	Relatively High	Relatively High
2007	12	Relatively High	Relatively High
2006	12	Relatively High	Relatively High
2005	12	Relatively High	Relatively High

الجدول ٦-١٣ وجود تشابهات كاملة بين الضغط الجوي المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-١٤، نلاحظ وجود شبه تشابهات بين الضغوط الجوية المرصودة والمتنبأ بها في بعض السنوات.

year	Month	Observed Pressure	Predicted Pressure
1996	1	High	Relatively High/High
2003	1	Relatively High	Relatively High/High
2002	1	Relatively High	Relatively High/High
2001	1	Relatively High	Relatively High/High
1998	1	Relatively High	Relatively High/High

الجدول ٦-١٤ وجود شبه تشابهات بين الضغوط الجوية المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-١٥، نلاحظ وجود فروقات واضحة بين الضغوط الجوية المرصودة والمتنبأ بها في بعض السنوات.

year	Month	Observed Pressure	Predicted Pressure
1994	1	High	Relatively High

1993	1	High	Relatively High
1991	1	High	Relatively High
1988	1	High	Relatively High
1987	1	High	Relatively High
1993	2	High	Relatively High
1992	12	Relatively High	High
2003	12	High	Relatively High

الجدول ٦-١٥ وجود فروقات واضحة بين الضغوط الجوية المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-١٦، نلاحظ وجود تشابهات كاملة بين ضغوط بخار الماء المرصودة والمتنبأ بها في بعض السنوات.

year	Month	Observed Vapor Pressure	Predicted Vapor Pressure
1986	1	High	High
2004	1	High	High
2007	1	High	High
1997	2	Relatively High	Relatively High
1996	2	Relatively High	Relatively High
1995	2	Relatively High	Relatively High
1994	2	Relatively High	Relatively High
1992	2	Relatively High	Relatively High
1991	2	Relatively High	Relatively High
1990	2	Relatively High	Relatively High
1989	2	Relatively High	Relatively High
1988	2	Relatively High	Relatively High
1987	2	Relatively High	Relatively High
1986	2	Relatively High	Relatively High
1985	2	Relatively High	Relatively High
2004	2	Relatively High	Relatively High
2003	2	Relatively High	Relatively High
2002	2	Relatively High	Relatively High
2001	2	Relatively High	Relatively High
2000	2	Relatively High	Relatively High
1999	2	Relatively High	Relatively High
1998	2	Relatively High	Relatively High
2007	2	Relatively High	Relatively High
2006	2	Relatively High	Relatively High
2005	2	Relatively High	Relatively High
1997	12	Very High	Very High
1995	12	Very High	Very High
1994	12	Very High	Very High
1993	12	Very High	Very High

1991	12	Very High	Very High
1988	12	Very High	Very High
1985	12	Very High	Very High
2001	12	Very High	Very High
2000	12	Very High	Very High
1998	12	Very High	Very High
2006	12	Very High	Very High

الجدول ٦-١٦ وجود تشابهات كاملة بين ضغوط بخار الماء المرصودة والمتنبأ بها في بعض السنوات

كما هو موضح في الجدول ٦-١٧، نلاحظ وجود شبه تشابهات بين ضغوط بخار الماء المرصودة والمتنبأ بها في بعض السنوات.

year	Month	Observed Vapor Pressure	Predicted Vapor Pressure
1997	1	Moderate	Moderate/High
1995	1	High	Moderate/High
1994	1	High	Moderate/High
1993	1	Moderate	Moderate/High
1992	1	High	Moderate/High
1990	1	High	Moderate/High
1989	1	High	Moderate/High
1988	1	High	Moderate/High
1987	1	Moderate	Moderate/High
2003	1	High	Moderate/High
2002	1	High	Moderate/High
2001	1	High	Moderate/High
2000	1	High	Moderate/High
1999	1	High	Moderate/High
1998	1	Moderate	Moderate/High
2006	1	High	Moderate/High
2005	1	High	Moderate/High
1989	12	Very High	High/Very High
1987	12	Very High	High/Very High
1986	12	High	High/Very High
2004	12	Very High	High/Very High
2003	12	Very High	High/Very High
1999	12	Very High	High/Very High
2007	12	High	High/Very High
2005	12	Very High	High/Very High

الجدول ٦-١٧ وجود شبه تشابهات بين ضغوط بخار الماء المرصودة والمتنبأ بها في بعض السنوات.

كما هو موضح في الجدول ٦-١٨، نلاحظ وجود فروقات واضحة بين الضغوط بخار الماء المرصودة والمتنبأ بها في بعض السنوات.

Year	Month	Observed Vapor Pressure	Predicted Vapor Pressure
1996	1	Moderate	High
1991	1	High	Moderate
1985	1	Moderate	High
1996	2	Moderate	High
1994	2	High	Moderate
2000	2	Moderate	High
1992	12	High	Very High
1990	12	High	Very High
2002	12	High	Very High

الجدول ٦–١٨ وجود فروقات واضحة بين الضغوط بخار الماء المرصودة والمتنبأ بها في بعض السنوات

عموماً من الجداول أعلاه، يمكننا أن نستنتج ما يلي :

- يتنبأ نظامنا بقيم مجالية. تكون بعض من هذه القيم متشابهة تماماً مع القيم المرصودة في بعض السنوات، وتكون بعض القيم الأخرى متشابهة جزئياً مع القيم المرصودة في سنوات أخرى، وتكون بعض القيم الأخرى مختلفة تماماً عن القيم المرصودة في السنوات الأخرى.
- تكون النسب المئوية للتشابهات التامة (الكاملة) ٣٦.٢٣، ٣٧.٦٨، ٢٤.٦٣، ٨١.١٦، ٥٢.١٧ بالمئة في حالة التنبؤ بدرجات الحرارة وكميات السحب والرطوبة النسبية وسرع الرياح والضغوط الجوية وضغوط بخار الماء، على التوالي.
- تكون النسب المئوية لأنصاف التشابهات ٥٧.٩٧، ٥٠.٧٢، ٤٠.٥٨، ٣٦.٢٦ ٧.٢٥، ٣٦.٢٣ بالمئة في حالة التنبؤ بدرجات الحرارة وكميات السحب والرطوبة النسبية وسرع الرياح والضغوط الجوية وضغوط بجار الماء، على التوالي.
- تكون النسب المئوية للفروقات (الاختلافات) الواضحة ٥.٧٩، ١١.٥٩، ٣٤.٧٨، ٣٧.٦٨، ١١.٥٩، ١٣.٠٤ بالمئة في حالة التنبؤ بدرجات الحرارة وكميات السحب والرطوبة النسبية وسرعة الرياح والضغوط الجوية وضغوط بخار الماء، على التوالي.

٦ – ٣ ملاحظات ختامية

في هذا الفصل، اختبرنا نظامنا الخبير الاستقرائي من ١٩٨٥ إلى ٢٠٠٧، في تطبيق مناخي. من مهمة الاختبار هذه، يمكن أن نستنتج ما يلي :

١– إذا اعتبرنا شبه التشابهات نجاح للنظام ، عندئذ تكون النسبة المئوية لنجاح النظام ٨١.٠٥ بالمئة بالنسبة إلى تطبيقنا المناخي.

٢– تكون النسبة المئوية لفشل النظام ١٩.٠٨ بالمائة بالنسبة لتطبيقنا المناخي.

الفصل السابع
النتائج والاتجاهات المستقبلية

٧ – ١ مقدمة

يحاول هذا الفصل أن يرسم النتائج بخصوص العمل المنجز، الأهداف الرئيسية والإنجازات، والحالة المستخدمة. وينظم هذا الفصل كما يلي : يعرض المقطع ٧-٢ الإسهامات الرئيسية لرسالتنا. ويناقش المقطع ٧-٣ السمات الرئيسية في النظام الخبير الاستقرائي المطور. وأخيراً، يختتم الفصل ببعض الاقتراحات البحث المستقبلي في المقطع ٧-٤.

٧ – ٢ إسهامات الأطروحة

كانت الأهداف الرئيسية في دراستنا التغلب على مشاكل اكتساب المعرفة من خبير النطاق (خبير المجال) بأتمتة عملية اكتساب المعرفة بتطوير منهج جديد يعتمد على مفاهيم وأساليب من تعلم الآلة والتنقيب بالمعطيات ونموذج تحليل الانحدار. ومن ثم تطبيق هذا المنهج على مجموعات معطيات الطقس لاكتساب المعرفة آلياً لتطوير نظام خبير استقرائي للتنبؤ بالطقس المدن في الجمهورية العربية السورية.

في إنجاز هدفنا :

١- طورنا خوارزمية تدعى " استقراء نموذج تحليل الانحدار الموجه بالصفة" لاكتشاف أنماط الطقس العامة في قواعد المعطيات الطقسية العلائقية الكبيرة.

٢- اقترحنا خوارزمية تدعى "توليد قواعد القرار" لتوليد قواعد التنبؤ بالطقس من مجموعة معطاة من عينات التدريب.

٣- كجزء من دراستنا لدور نظرية المجموعات التقريبية في تقييم الفائدة المحتملة للقواعد المكتشفة، اقترحنا قياس لعامل التأكد (المصداقية) لكل قاعدة من القواعد المولدة. ويكون القياس المقترح معتمداً على دالة العضوية التقريبية ويمكننا من تخصيص (تحديد) الأوليات النسبية لهذه القواعد والتي يمكن أن تستخدم في ترتيب مجموعة قواعد القرار الاحتمالية المولدة.

٤- لاختبار وتجريب الخوارزميات تعلم قاعدة المعطيات العلائقية المطورة، بنينا نظام تعلم قاعدة معطيات تجريبي أطلقنا عليه DBROUGH لاكتشاف المعرفة من المعطيات للتنبؤ بالطقس.

٥- كجزء من اختبار نظام DBROUGH المطور، أجرينا بعض التجارب الهامة لاختبار النظام المنفذ في تطبيقين مناخيين حقيقيين، وسردنا بعض النتائج التجريبية، ومن ثم عرضنا تقييم القواعد المكتشفة بواسطة خبير أرصاد جوية. تحدد كل قواعد المعرفة المكتشفة ما يعتقده الخبير على أنه علاقة ارتباط منطقية في مجال الدراسة، ونتيجة لذالك، قبلت من قبل خبير النطاق.

٦- لتطبيق القواعد المكتشفة في التنبؤ بالطقس، بنينا نظام خبير استقرائي فريد للتنبؤ بالطقس، أطلقنا عليه IESP، بحيث أن نظام DBROUGH أصبح كنظام فرعي (جزئي) منه.

٧- كجزء من عملية اختبار نظام IESP المطور، اختبرنا النظام لمدة ٢٣ سنة من ١٩٨٥ حتى ٢٠٠٧ في تطبيق مناخي. وكانت نسبة نجاح النظام ٨١,٠٥ %. وكانت النسبة المئوية لفشل النظام ١٩.٠٨ %.

٧ – ٣ السمات الهامة IESP

في هذا المقطع، لخصنا ما يمكن اعتباره بالسمات الأهم لنظام IESP المقترح. ويمكن إيجاز هذه السمات بما يلي :

١. يحوي المعرفة في الشكل القواعد التي تساعد المتنبئ في التنبؤ بالطقس.
٢. يملك ٦ قواعد معرفة.
٣. يملك مجموعتي تدريب و٦ مجموعات اختبار.
٤. يفصل المعرفة عن التحكم.
٥. يملك آلية استدلال (محرك الاستدلال) يستخدم التسلسل المتقدم على طول خليط تتابعي ومتوازي من عوامل الثقة.
٦. يملك نظام اكتساب معرفة فرعي (DBROUGH) لاكتشاف أنماط الطقس العامة والقواعد من قواعد المعطيات الطقسية العلاقية (أي، مجموعات التدريب والاختبار). ويملك هذا النظام الفرعي عدة سمات: (١) يستقرئ نظام معلوماته بتطبيق خوارزمية الاستقراء الموجه بالصفة. (٢) يستطيع أن يحدد المفاهيم في نظام المعلومات لصفة قرار معينة بمعرفة المعلومات المتاحة. (٣) يستطيع أن يحدد المجموعة الفرعية الأساسية للصفات غير مكررة (غير زائدة) التي تعين مهمة الاكتشاف لصفة قرار معينة. هذا يعني، يستطيع النظام أن يستخرج السمات الملائمة من مجموعة محددة من السمات. (٤) يستطيع أن يولد قواعد المعرفة مع عوامل التأكد لمفهوم معين بمعرفة المعلومات المتاحة من مجموعة معطيات الهدف.
٧. يسمح لمعرفته لأن تعدل بسهولة بتنقيح أو بحذف أو بإضافة القواعد.
٨. يملك سطح بيني للنظام قوي ومرن، متضمناً السطح البيني للمستخدم والسطح البيني للمطور.
٩. يملك نظام توضيح فرعي (مبرر) يقدم شرح وتبرير لأفعال النظام (الاستفسارات HOW وWHY).

٧ – ٤ الاتجاهات المستقبلية

بعيداً عن المواضيع الخاصة المعالجة في العمل الحالي، يوجد بعض المواضيع الأخرى التي تفتح الأبواب على عدة مواضيع هامة للبحث المستقبلي. ويمكن أن نشير إلى الأهم من هذه المواضيع بما يلي :

١. قدمت هرميات المفهوم في دراستنا من قبل خبير نطاق. ويمكن أن يؤثر هذا على نتائج تحليل المجموعات التقريبية الموجهة بالصفة والتي عندئذ يمكن أن تستلزم نوعاً ما من عدم التأكد (عدم المصداقية) في النتائج التي يجب أن تؤخذ بعين الاعتبار أثناء التحليل. وبناءً على ذلك، يجب أن تركز دراستنا المستقبلية على التوليد الآلي (الأوتوماتيكي) لهرميات المفهوم، بحيث يكون النظام قادراً لأن يكتشفها أو يولدها ألياً اعتماداً على توزيع معطياتها في علاقة المعطيات الابتدائية (قاعدة المعطيات).
٢. من المعروف أن قواعد المعطيات ديناميكية. ويكون إعادة توليد المختزلات والقواعد لكل بديل مهمة غالية وشاقة. وبناءً على ذلك، فإنه من المفيد جداً من وجهة نظر الكفاءة (الفعالية) أن نكون قادرين لأن نعدل المختزلات والقواعد الموجودة في شكل تزايدي بدلاً من إعادة توليدها.
٣. دراسة موضوع تعميم القواعد، وبالتالي تمديد النظام لكي يكون قادراً على توليد القواعد المعممة أعظمياً أيضاً. وهذا يعني، القواعد التي تحتوي على عدد أصغر من الشروط. ومن ثم اختبار النظام وفقاً لهذه القواعد ومقارنة النتائج المحصول عليها مع النتائج المحصول عليها اعتماداً على قواعد القرار الاحتمالية.
٤. تمديد النظام ليشمل (ليتضمن) كل عناصر الطقس.
٥. تحسين السطح البيني للنظام (واجهة النظام) باستخدام الوسائط المتعددة وأساليب معالجة اللغات الطبيعية.

المراجع

Refrences:

1. Bulent KISKAC and Harun yARDIMCI Computer Engineering: Weather Prediction Expert System Approaches. (Ceng-568 Literature Survey) Middle East Technical University Ankara,TURKEy June, 2004.
2. J. O. Ayoade. Introduction to Climatology For The Tropics, John Wiley & Sons, 2003.
3. Ahrens, Essentials of Meteorology, West Publishing Company, 1993.
4. y. S. Takroni, Exploiting Rough Set Theory for Discovering Rules from Weather Database for Developing Weather Forecasting System, 10th International Conference on Artificial Intelligence Applications, Homs International Conference Center, Homs, Egypt, Feb. 6-Feb. 9, 2002.
5. Pristinsky, Prediction-Methods, Techniques, and Applications, Lappeenranta University of Technology, Information Technology Department, 2000.
6. J. Conway, Expert Systems and Weather Forecasting, Meteorological Magazine, 118,23-30, 1989.
7. K. Hansen, Weather Prediction Using Case-Based Reasoning and Fuzzy Set Theory, Dalhousie University-Daltech, 2000.
8. D. Meyer, A Hybrid Model for Predicting Fog and Stratus Clouds, Preprints of the 6th Conference on Aviation Weather Systems, American Meteorological Society, 441-444, 1995.
9. F. R. Mosher, Meteorological Craftsman US. Meteorological Engineers, Preprints of the 1st Conference on Artificial Intelligence, American Meteorological Society, 164-167, 1998.
10. Christopherson, Artificial Intelligence in the Weather Forecast Office-One Forecaster's View, Preprints of the 1st Conference on Artificial Intelligence, American Meteorological Society, 136-143, 1998. J
11. R. Kumar, Toward Building an Expert system for Weather Forecasting Operations, Expert System with Applications, Vol. 7, No. 2, April-June, P 373-381, 1994.
12. Durkin, Expert System Design and Development, Macmillan, Canada, Inc., 1994.
13. J. R. Quinln, Discovering rules by induction from large collection of examples, In D. Michie (ed.), Expert systems in micro electronic age. Edinburg University Press, 169-201, 1999.
14. C. Owen, Forecaster: An Expert System in Weather Forecasting, Proceedings of the 1st Florida Artificial Intelligence Research Symposium, April 1988.
15. Quinlan, J. R. C4. 5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

16. S. B. Kotsiantis, Supervised Machine Learning: A Review of Classification Techniques, Informatica 31(2007) 249-268, 2007
17. J. R. Quinlan. Improved use of continuous attributes in c4. 5. Journal of Artificial Intelligence Research, 4:77-90, 1996.
18. J. R. Quinlan, C4. 5 Programs for Machine Learning, Morgan Kaufmann Publishers, Inc., 1993.
19. P. Joshi, Analysis of Data Mining Algorithms, 1997.
20. J. R. Quinlan, C5. 0: An Informal Tutorial, 2000.
21. Agrawal, A. Arning, T. Bollinger, M. Mehta, j. Shafer, and R. Srikant, The Quest Data Mining System, proc. Of the 2 nd international Conference on Knowledge Discovery in Databases and Data Mining, Portland, Oregon, August 1996.
22. Matthew N. Anyanwu Sajjan G. Shiva, Comparative Analysis of Serial Decision Tree Classification Algorithms.
23. J. Han and M. Kamber, Data Mining Concepts and Techniques, Morgan Kaufmann Publishers, 2001.
24. Z. Pawlak, J. Grzymala-Busse, R. Slowinski, and W. Ziarko, Rough Sets in Communication of the ACM. Vol. 38, No. 11, November 1995.
25. Z. Pawlak, Rough Sets in International Journal of Computer and Information Sciences, 11, pp. 341-356, 2002.
26. Pawlak, Hard and Soft Sets in Rough Sets, Fuzzy Sets and Knowledge Discovery, Britian Computer Society, Great Britain, 1994.
27. Turban, Efraim, and Mclean, Ephraim, and Wetherbe, James, 2002, Information Technology for Management,3rd Ed., (Danvers, MA: John Wiley & Sons).
28. Turban, Efraim, and Aronson, Jay E., 2001, DECISION SUPPORT SySTEMS and INTELLIGENT SySTEMS, 6th ed., (New Jersey: Prentice-Hall).
29. Post, Gerald V., and Anderson David L., 2000, Management Information Systems, 2nd Ed., (USA: Mc Graw-Hill).
30. Michie, S. Muggleton, C. Riese, and S. Zubrick, RULEMASTER: A second-generation knowledge- engineering facility, Proceedings of the First Conference on Artificial Intelligence Applications, IEEE Computer Society, Dec. 2004.
31. N. Rodionov and J. H. Martin, An Expert system-Based Approach to Prediction of year-to year Climate Variations in the North Atlantic Region, University of Colorado, Computer Science Department, 1998.
32. J. Han, y. Cai, N. Cercone, and y. Huang, Discovery of Data Evalution Regularities in Large Databases, the School of Computing Science, Simson Fraser University, British Columbia, Canada, 1993.
33. Lu and J. Han, Discovery of General Knowledge in Large Spatial Databases, the School of Computing Science, Simson Fraser University, British Columbia, Canada, 1993.
34. P. S. Mann. Introductory Statistics, John Wiley & Sons, Inc., 1995.

35. M. Saadie, Regression and Correlation Analysis, Faculty of Commerce –Ain shams University, 1993.
36. H. Sanders. Statistics: A First Course, McGraw-Hill, Inc., 1995
37. W. Pedrycz and F. Gomide, An Introduction to Fuzzy Sets Analysis and Design, Massachusetts, Institute of Technology, 1998.
38. y. S. Takroni and A. H. Al-Omar, Discovery of General Weather Patterns in Large Relational Weather Databases, International Journal of Applied Mathematics, Volume 7 No. 1, Sofia, Bulgaria, 2001.
39. H
40. y. Zhongjia, Translator: Zenglingsen. Meteorology, China Science and Technology Press 32 Baishiqio road, Beijing, China, 1985.
41. W. Ziarko, Variable Precision Rough Set Model. Journal of Computer and System Sciences, Vol. 46, No. 1, 1993, pp. 39-59.
42. Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, 1991.
43. Cai, N. Cercone and J. Han, Attribute-Oriented Induction in relational Databases, in G. Piatetsky Shapiro and W. J. Frawlly (eds.), Knowledge Discovery in Databases, AAAI/MIT Press, pp. 213-228, 1991.
44. W. Ziarko, N. Shan, A Rough Set-Based Method for Computing All Minimal Deterministic Rules in Attribute-Value Systems, Computational Intelligence: An International Journal, to appear, U. of R. Technology Report CS-93-02.
45. W. Ziarko, N. Shan, Data-Based Acquisition and Incremental Modification of classification rules. An International Journal, Volume 11, Issue2, May, pp. 357-369, 1995.
46. Shan, x. Hu, W. Ziarko, N. Cercone, A Generalized Rough Sets Model, Proceedings of the third Pacific Rim International Conference on Artificial Intelligence, China. 1994, pp. 437-443.
47. Hu, Knowledge Discovery in Databases: An Attribute-Oriented Rough Set Approach, Ph. D. Thesis, Faculty of Graduate Studies, University of Regina, Regina, Saskatchewan, June, 1995.
48. J. Frawley, G. Piatetsky-Shapiro and C. J. Matheus, Knowledge Discovery in Databases: An overview, in G. Piatetsky-Shapiro and W. J. Frawley (eds.), Knowledge Discovery in Databases, AAAI/MIT, Press, 1991, pp. 1-27.
49. H. Al-Omar, Discovering Rules From Data for Weather Forecasting Using Rough Set Theory, International Journal of Applied Mathematics, Volume 6 No. 1, Sofia, Bulgaria, 2001.
50. P. M. Tag, R. L. Bankert, M. Hadjimichael, A. P. Kuciauskas, W. T. Thompson, and K. L. Richardson, Applying Knowledge Discovery from Databases (KDD) to Combined Satellite and High Resolution Numerical Model Data, Proceedings AMS Second Conference on Artificial Intelligence, Long Beach, CA, Jan 2000.
51. M. Kryszkiewiez, Maintenance of Reducts in the Variable Precision Rough Sets Model, Institute of Computer Science, Warsaw University of Technology, Nowowiejska 15/19, 1994.

52. T. Aasheim and H. G. Solheim, Rough Sets as a Framework for Data Mining, Knowledge Systems Group, Faculty of Computer Systems and Telematics, the Norwegian University of Science and Technology, Trondbeim, 30th April 1996.
53. J. Katzberg and W. Ziarko, Variable Precision Rough Sets with Asymmetric Bounds, In Ziarko, W. (ed.) Rough Sets, Fuzzy Sets and Knowledge Discovery, Springer-Verlag, 1994, pp. 167-177.
54. W. Ziarko, Variable Precision Rough Set Model. Journal of Computer and System Sciences, Vol. 46, No. 1, 1993, pp. 39-59.
55. P. Kuciauskas, L. R. Brody, M. Hadjimichael, R. L. Bankert, P. M. Tag, and J. E. Peak, A Fuzzy Expert System to Assist in the Prediction of Hazardous Wind Conditions in the Mediterranean Basin, Meteorological Applications (Royal Meteorological Society), 5, 307-320, 1998.
56. Elio and J. De Haan, Representing quantitative and qualitative knowledge in a knowledge-based storm-forecasting system, International Journal of Man-Machine Studies, Vol. 25, No. 5, PP. 523-547, Novermber 1986.

٥٧. تقنيات التنقيب في قواعد البيانات واستكشاف المعلومات المخبأة فيها إعداد أ. مصطفى فؤاد عبيد محلل وخبير متخصص في الشئون المعلوماتية – فلسطين.

٥٨. المناخ د. إبراهيم محمد على بدوى.

الجداول

الأشكال

الاختصارات

رمز الاختصار		رقم الصفحة
MSE	خدمة الأرصاد الجوية	٥
AI	الذكاء الصنعي	٥
NWF	مخرجات النماذج العددية	٦
KDD	اكتشاف المعرفة في قواعد البيانات	٢٦
SSE	مجموع البواقي للمربعات	٤٤
SSY	المجموع الكلي للمربعات	٤٤
CFS	عامل التأكد	٤٧
AS	نظام معلومات	٥٧
RED	المختزل	٦٠
DBRough	نظام تعلم قاعدة المعطيات	٦٣
GUI	سطح بيني رسومي	٦٤
NC	عينة تدريب	٧٤
IESP	نظام خبير استقرائي للتنبؤ بالطقس	٨٠

المصطلحات

انكليزي – عربي

A

Approximation Space	فضاء التقريب
Artificial Intelligent	الذكاء الصنعي
Attribute-Oriented Induction	الاستقراء الموجه بالصفة

B

Boundary Region	المنطقة الحدية

C

Cluster Analysis	التحليل العنقودي
Concept	المفهوم
Condition Attributes	الصفات الشرطية
Control Mechanism	آلية التحكم

D

Decision Attribute	صفة القرار
Decision Tree	شجرة القرار

E

End-User	المستخدم النهائي
Expert	الخبير

Expert System	النظام الخبير

G

Generalization	التعميم

I

Indiscernibility Relation	علاقة عدم إمكانية التمييز
Induction	الاستقراء
Information System	نظام المعلومات
Interface Engine	محرك الاستدلال

K

Knowledge Base	قاعدة المعرفة
Knowledge Engineer	مهندس المعرفة

L

Lower Approximation	التقريب السفلي

M

Multiple Regression Equation	الانحدار الخطي المتعدد

N

Nearest Neighbor	الجار الأقرب
Negative Region	المنطقة الموجبة
Neural Network	الشبكات العصبية

P

positive region	المنطقة السالبة
Probability approximation space	فضاء التقريب الاحتمالي
Programmer	المبرمج

R

Reduction	المختزل
Reduction of condition attributes	اختزال الصفات الشرطية
Regression Analysis Model	نموذج تحليل الانحدار
Rough Set Theory	نظرية المجموعات التقريبية

S

Simple Regression Equation	الانحدار الخطي البسيط

U

Universe	المجموعة الشاملة
upper approximations	التقريب العلوي
User Interface	واجهة المستخدم

الفهرس

الفصل الأول

١. مقدمة

الفصل الثاني

٢. دراسة مرجعية

الفصل الثالث

٣-. اكتشاف أنماط الطقس العامة من قواعد المعطيات العلائقية الكبيرة

الفصل الرابع

٤-. اكتشاف القواعد من المعطيات للتنبؤ بالطقس باستخدام نظرية المجموعات التقريبية (تصميم و تنفيذ و تجريب الخوارزميات)

الفصل الخامس

٥-. تنفيذ نظام خبير استقرائي للتنبؤ بالطقس

الفصل السادس

✍-٦. اختبار النظام

الفصل السابع

✍-٧. النتائج والاتجاهات المستقبلية

الملحق A

في هذا الملحق، نقدم التحاور بين نظام IESP والمستخدم والذي يوضح تشغيل النظام.

ASTRO Login Interface

Using Inductive Expert Systems in Prediction
Using
Attribute_Oriented Rough Set Theory

IESP System Vresion 1.0

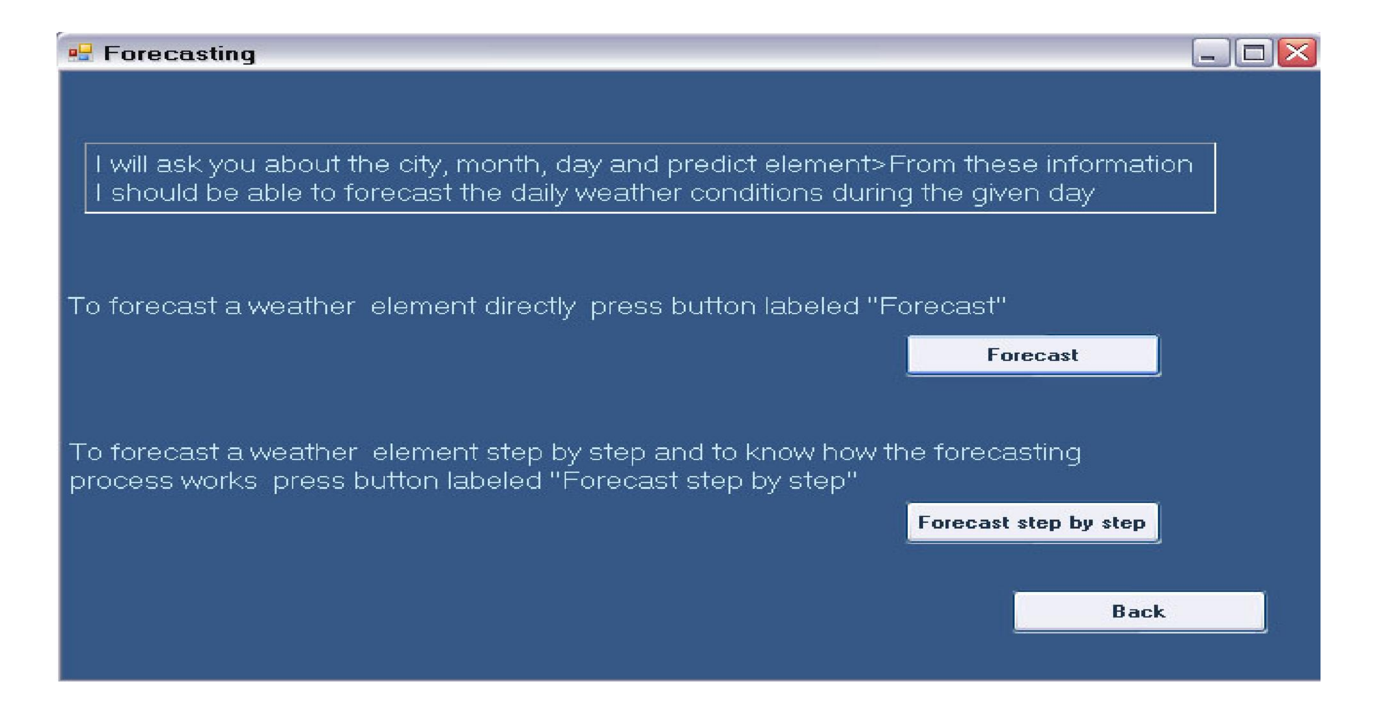
If You would likt to inter to forecasting system write "user" else If You would like to enter the DBROUGH,Training and test sets write "expert"

User Name user
Password *****
Login Exit

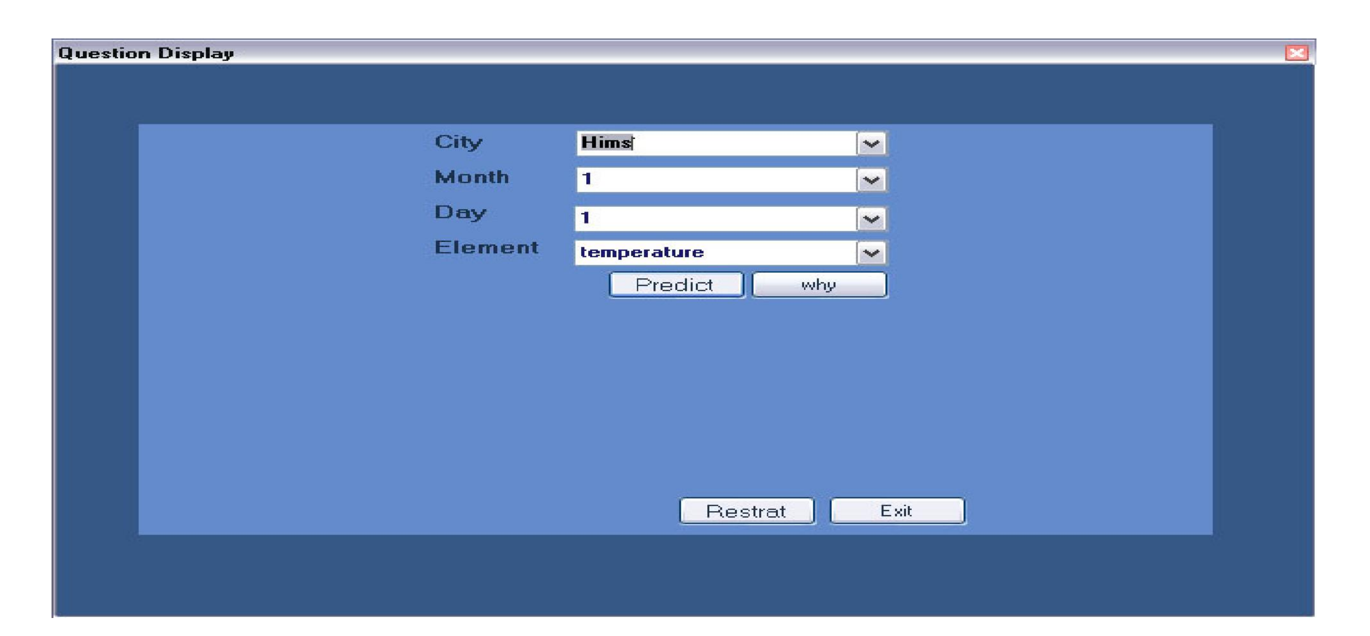
الشاشة الأولى هي واجهة من أجل تسجيل الدخول لاختيار الدخول كمستخدم أو كخبير، في حـال اختيـار الدخول كمستخدم يكتب "user" من أجل الدخول إلى النظام الخبير. فتظهر الشاشة التالية:



تخبر هذه الشاشة المستخدم أنه تم دخولـه إلـى النظـام الاسـتقرائي، وباختيـار الـزر "Forecasting Knowledge Based System" تظهر الشاشة التالية:

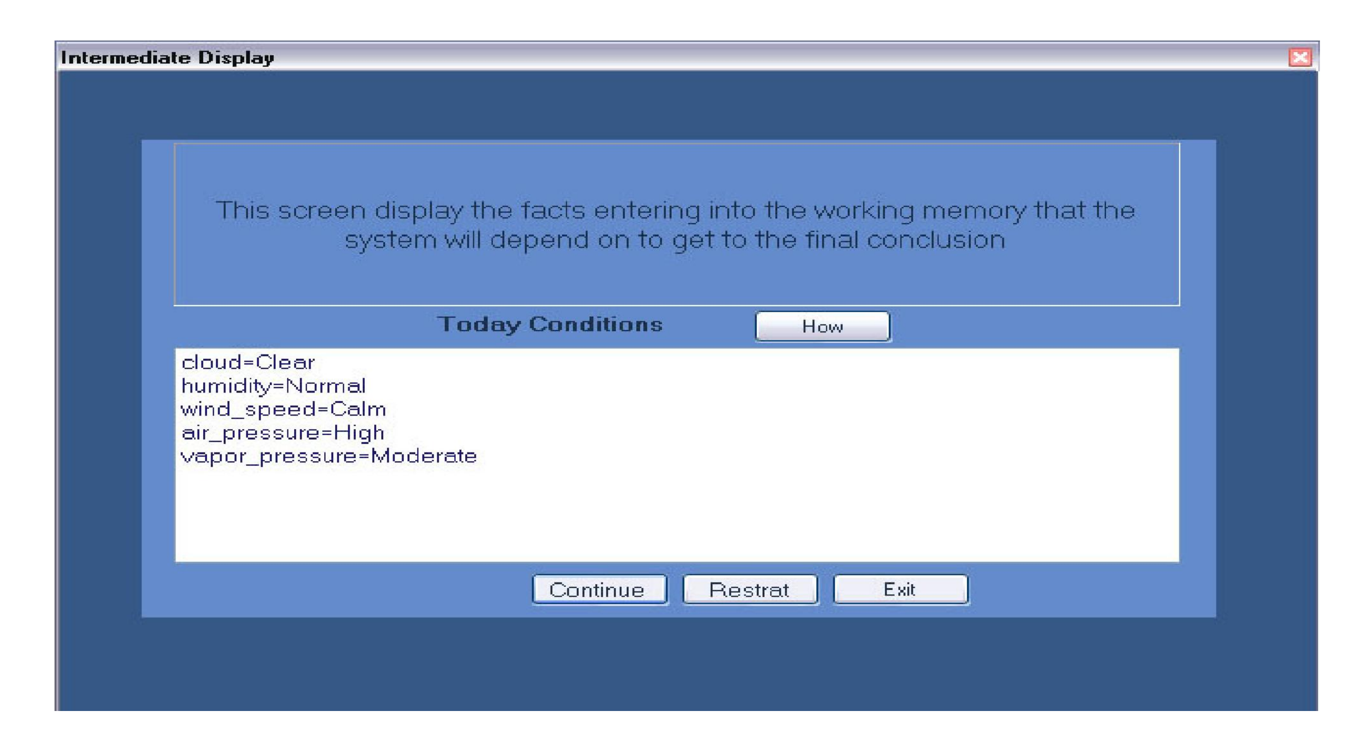


الشاشة السابقة تخبر المستخدم أنه سوف يتم طرح بعض الأسئلة عليه، هذه الأسئلة التي سيستخدمها النظام للحصول على النتيجة النهائية. كما تعطي هذه الشاشة الخيار للمستخدم إما أن يختار التنبؤ المباشر أو أن تتم عملية التنبؤ خطوة خطوة، أي يتم شرح الخطوات التي مر بها النظام للحصول على نتيجة التنبؤ. وفي حال اختيار الخيار الثاني تظهر شاشة الأسئلة التالية:

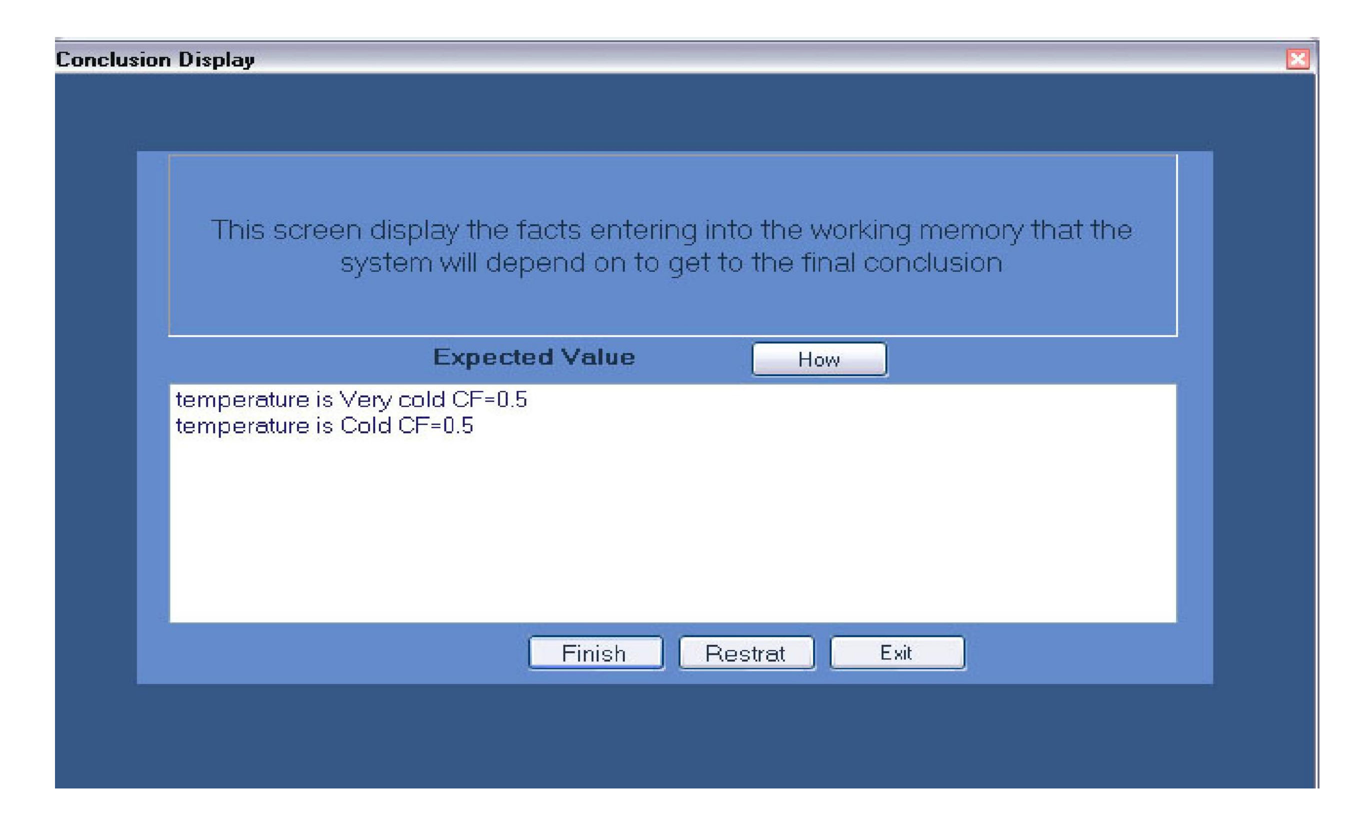


في الشاشة السابقة، يطلب من المستخدم تقديم الإجابة على الأسئلة الخاصة بالمنطقة والشهر واليوم وعامل الطقس المراد إجراء عملية التنبؤ له. وبفرض أن المستخدم قد حدد أن المدينة هي حمص، والشهر هو كانون الثاني، واليوم هو اليوم الأول، وعنصر الطقس المراد هو درجة الحرارة، عندئذ النظام يستخدم هذه الحقائق وبالتالي تظهر شاشة النتائج المرحلية "Intermediate Display" التالية:

يستخدم النظام الحقائق التي وصل إليها في الشاشة السابقة ضمن القواعد الاحتمالية التي وصل إليها للحصول على النتيجة النهائية.



في الشاشة المرحلية الأولى تظهر في هذه الشاشة نماذج الطقس العامة التي وصل إليها النظام وبالضغط على الزر "Continue" تظهر الشاشة النتائج الأخيرة "Conclusion Display" التالية:



يستخدم النظام الحقائق التي وصل إليها في الشاشة السابقة ضمن القواعد الاحتمالية التي وصل إليها للحصول على النتيجة النهائية.